SYMMETRY-ADAPTED PERTURBATION THEORY FOR ORGANOCATALYSIS

A Thesis
Presented to
The Academic Faculty

by

Brandon W. Bakr

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Chemistry and Biochemistry

Georgia Institute of Technology
August 2018

SYMMETRY-ADAPTED PERTURBATION THEORY FOR ORGANOCATALYSIS

Approved by:

Professor C. David Sherrill, Advisor
School of Chemistry and Biochemistry
Georgia Institute of Technology

Professor Thomas Orlando
School of Chemistry and Biochemistry
Georgia Institute of Technology

Professor Joseph W. Perry
School of Chemistry and Biochemistry
Georgia Institute of Technology

Professor Jesse McDaniel
School of Chemistry and Biochemistry
Georgia Institute of Technology

Professor Edmond Chow
School of Computational Science and Engineering
Georgia Institute of Technology

Date Approved: 26 April 2018

This thesis is dedicated to my loving parents, A.D. and Merlene Bakr, whose love and support made this endeavor possible.

ACKNOWLEDGEMENTS

I would like to thank my doctoral advisor, Prof. C. David Sherrill, for being an exemplary scientist during my whole time at Georgia Tech. His ability to pick out important points of scientific stories and craft interesting and important stories from scientific results has been unparalleled in my academic career in computational chemistry. Additionally, his support and guidance greatly assisted me during my graduate studies and enabled me to finish my doctorate. He takes great care in crafting the quality of his group and ensures that everyone is on the right path by communicating with members. I am very appreciative of being able to be involved with the highly skilled and diverse group of researchers that he has assembled. Additionally, I am very thankful for the external opportunities that he as allowed and helped me pursue. Without him, my internship at OpenEye and numerous trips to meet a number of researchers at various places across the country would not have been possible.

I would also like to thank those who taught me to love learning and introduced me to chemistry. Mrs. Judy Thomas, my third grade teacher, was very influential in my early schooling and ensured that I would adopt a penchant for learning and mathematics. Also, my chemistry teacher in high school, Mrs. Valencia Snell, did an extraordinary job of introducing me to chemistry and gave me the tools to pursue the subject at the college level. Additionally, my mentors at Auburn University equipped me with the tools to make it through graduate school. Professor Andras Bezdek mentored me in mathematics research and taught me to always approach difficult tasks with confidence. I have no doubt that I would have made it this far in my academic career without this lesson. Professor Konrad Patkowski provided me with an extremely beneficial and productive research experience during the final year of my undergraduate career. Without the exposure to computational chemistry that he provided, I may never have pursued or been able to pursue this field in my graduate studies. Professor John Gorden also greatly assisted me in learning to enjoy

and respect chemistry. His mentorship and energetic conversations always inspired me to continue my pursuit of chemistry.

While in Professor Konrad Patkowski's lab at Auburn University, I had the pleasure of working with and being mentored by Dr. Daniel G. A. Smith. His mentorship has proven invaluable in both my undergraduate and graduate studies. He did a great job of introducing me to Python, currently my favorite programming language, when I first started learning computational chemistry. Dr. Rob Parrish was instrumental during my start in the Sherrill group and very helpful when I was learning local MP2. Dr. Lori A. Burns was a great co-worker that always could assist me with my Psi4 problems and questions. Dr. Jérôme Gonthier was a great co-worker and friend who had an ability to brighten a day with his happy attitude. Trent Parker was an inspirational co-worker and friend during my time at Georgia Tech and was always very helpful when I asked about random things from new buildings in Atlanta to the economics of farming. I am very thankful to have become co-workers and friends with Dr. Carlos Borca, Dr. Ryan Richard, Matt Schieber, and Mike Zott. These gentlemen were very knowledgeable of work-related topics and life in general, and they made my time at Georgia Tech fun and more productive. I would also like to thank the Sherrill Group members that were very inspiring during my career at Georgia Tech: Dr. Matt Kennedy, Yi Xie, Dom Sirianni, Constance Warden, Asim Alenaizan, and Dr. Daniel Nascimento.

I would like to thank Dr. Tom Darden for his encouragement and lessons while at OpenEye Scientific Inc. Dr. Darden and I chatted a lot about science and life during my time with OpenEye, and he was a great source of encouragement for continuing my graduate studies. Dr. Anthony Nicholls was very gracious to support me while at OpenEye, for which I am very appreciative. Dr. Grigory Ovanesyan and David Rower, of OpenEye, also provided unending encouraging and insightful discussions. David has pushed me to new heights both literally and metaphorically, as he was my hiking buddy in New Mexico. Lastly, for external collaborators, I would like to thank Dr. Dan Cheney, Dr. Doree Sitkoff, and Dr. Sean Zhu for their encouragement and hosting me during my visit to Bristol-Myers Squibb.

Finally, I would like to thank my parents, A.D. and Merlene Bakr. My father is a very hard working man and has taught me to do whatever it takes to get a job done. My mother has unfathomable belief in my capabilities and has encouraged me to push through challenges even in the darkest of times. Without their great parenting and support, I would not have a chance of making it to where I am today.

TABLE OF CONTENTS

DEDICATION . . . . . **iii**

ACKNOWLEDGEMENTS . . . . . **iv**

LIST OF TABLES . . . . . **x**

LIST OF FIGURES . . . . . **xi**

SUMMARY . . . . . **xvi**

I INTRODUCTION **1**

1.1 Brief Introduction to Electronic Structure Theory 1

1.1.1 Schrödinger Equation 1

1.1.2 Hartree-Fock Theory 2

1.1.3 Symmetry-Adapted Perturbation Theory 4

1.2 Post Hartree-Fock Methods of Electron Correlation 7

1.2.1 Møller-Plesset Perturbation Theory 7

1.2.2 Local MP2 9

1.2.3 Coupled-Cluster Theory 12

1.3 Density Fitting 14

1.4 Organization of Thesis 15

II EXAMINATION OF NON-COVALENT INTERACTIONS IN THE HOUK-LIST MECHANISM **16**

2.1 Summary 16

2.2 Introduction 17

2.3 Computational Methods 20

2.4 Results and Discussion 23

2.4.1 $NCH^{\delta+} \cdots {}^{\delta-}O{=}C$ Interactions 24

2.4.2 C–H/π Interaction 29

2.4.3 Phenyl *ortho*-Hydrogen $\cdots$ Carboxylic Acid 31

2.4.4 Summary of Non-Covalent Interactions and Comparison to Experiment 33

2.5 Conclusions 37

III ANALYSIS OF TRANSITION STATE STABILIZATION BY NON-COVALENT INTERACTIONS IN THE ADDITION OF ORGANOBORON REAGENTS TO FLUOROKETONES 39

3.1 Summary 39

3.2 Introduction 39

3.3 Theoretical Methods 41

3.4 Results and Discussion 45

3.4.1 H $\cdots$ F 46

3.4.2 H $\cdots$ O 48

3.4.3 O $\cdots$ O 50

3.4.4 O $\cdots$ F 50

3.4.5 Steric Interactions in Reaction B 51

3.4.6 Steric Interactions in Reaction C 55

3.5 Conclusions 59

IV SUBSTITUTION OF LOCAL MP2 IN B2PLYP 61

4.1 Summary 61

4.2 Introduction 61

4.3 Theoretical Methods 62

4.3.1 DH-DFT Methods 62

4.3.2 Local Correlation Methods 63

4.3.3 Computations 64

4.4 Results and Discussion 64

4.5 Conclusions 71

V AUTOMATION OF WORKFLOWS INVOLVING F-ISAPT 74

5.1 Summary 74

5.2 Motivation 74

5.3 Solutions 75

5.3.1 Protein Data Bank 76

5.3.2 XYZ 78

5.3.3 PyMOL Interface 79

5.4 Conclusion 80

VI CONCLUSIONS AND OUTLOOK 82

APPENDIX A — EXAMINATION OF NON-COVALENT INTERACTIONS IN THE HOUK-LIST MECHANISM 86

APPENDIX B — ANALYSIS OF TRANSITION STATE STABILIZATION BY NON-COVALENT INTERACTIONS IN THE ADDITION OF ORGANOBORON REAGENTS TO FLUOROKETONES 97

REFERENCES 101

VITA 108

LIST OF TABLES

1 The relative total energies of the transition states (ΔE), relative Gibbs free energies at 298K ($\Delta\Delta G$), and percent populations are listed for the CCSD(T)/CBS estimates obtained in this work and the previous best estimate by Rzepa and co-workers[2] at the B3LYP-D3/TZVP/SCRF=DMSO level of theory. CCSD(T)/CBS ($\Delta\Delta G$) results include solvent and thermodynamic corrections obtained at the B3LYP-D3/TZVP/SCRF=DMSO level. Experimentally measured[3] product ratios are presented for comparison. . . 36

2 Computed interaction energies for B2PLYP variants in the cc-pVTZ basis and benchmark values are listed. 69

3 Computed interaction energies for B2PLYP variants in the cc-pVTZ basis and benchmark values are listed. 70

4 Mean absolute errors (MAEs) relative to benchmark values for S22 interaction energy computations in the cc-pVDZ basis are listed. 71

5 Mean absolute errors (MAEs) relative to benchmark values for S22 interaction energy computations in the cc-pVTZ basis are listed. 71

LIST OF FIGURES

1 The proline-catalyzed reaction between benzaldehyde and cyclohexanone along with the four possible stereochemical outcomes [(*S*,*R*) *anti*, (*S*,*S*) *syn*, (*R*,*R*) *ent-syn*, and (*R*,*S*) *ent-anti*]. Proline catalyzes the reaction by reacting with cyclohexanone to form a reactive enamine intermediate. 18

2 Transition states, without specified conformations, leading to each possible stereoisomer [(*S*,*R*) *anti*, (*S*,*S*) *syn*, (*R*,*R*) *ent-syn*, and (*R*,*S*) *ent-anti*] of the product. The stereochemistry is determined by the relative orientations of the phenyl group of benzaldehyde and the forming carbon-carbon bond. 19

3 The four low-lying conformations for each transition state stereoisomer. The pyrrolidinyl ring is puckered away from benzaldehyde in conformers 1 and 3, and puckered towards it in conformers 2 and 4. Cyclohexene adopts a chair conformation in conformers 1 and 2, and a boat conformation in conformers 3 and 4. 19

4 The (*S*,*S*) *syn* 1 transition state, with the three hypothesized sources of non-covalent stabilization: $NCH^{\delta+} \cdots {}^{\delta-}O{=}C$, C–H/$\pi$, and carboxylic acid $\cdots$ *ortho*-hydrogen 20

5 The fragmentation scheme used for all F-SAPT analyses in this work, unless otherwise stated. There are five distinct fragments in total for the two monomers. Enamine intermediate: 1) carboxylic acid, 2) cyclohexene, 3) pyrrolidinyl. Benzaldehyde: 4) aldehyde and 5) phenyl. 22

6 The electrostatic interaction between the benzaldehyde monomer and the entire enamine intermediate is visualized for (*S*,*R*) *anti* 1 (left) and (*R*,*R*) *ent-syn* 1 (right) using the terms from an F-SAPT0/jun-cc-pVDZ decomposition. Deeper red indicates a stronger attraction, and deeper blue indicates stronger repulsion. 25

7 The F-SAPT predicted energetic components of the interaction between each of the fragments are plotted at the F-SAPT0/jun-cc-pVDZ level of theory. Only the first two letters of the stereochemical designation along with the corresponding conformer number are used to represent each possible transition state. 26

8 The primary charges involved in determining the favorability of the interaction between $NCH^{\delta+}$ of the enamine intermediate and ${}^{\delta-}O{=}C$ of benzaldehyde for the (*S*,*R*) *anti* 1 transition state. 27

9 The dispersion interaction between the phenyl group and the entire enamine intermediate is visualized for (*S*,*S*) *syn* 1 (left) and (*S*,*R*) *anti* 1 (right) using our F-SAPT0/jun-cc-pVDZ decomposition. In this figure, aldehyde interactions are excluded. 29

10 The electrostatic interaction between the phenyl group and the enamine intermediate is visualized for (*S*,*S*) *syn* 1 (left) and (*R*,*R*) *ent-syn* 1 (right) at the F-SAPT0/jun-cc-pVDZ level of theory. In this figure, aldehyde interactions are excluded. 32

11 The total SAPT decomposition of the energetics in each transition state is plotted at the SAPT0/jun-cc-pVDZ level of theory. This plot only accounts for total monomer interactions, meaning the energetics are for the whole enamine intermediate interacting with benzaldehyde. 35

12 Enantioselective addition of an allyl group to fluoroketones, through a complex formed between the allyl-boron reagent and an aminophenol catalyst. . 40

13 Transition state structures for the reaction between trifluoro-methyl ketones and organoboron catalyst complexes from Figure 12, using the numbering from Hoveyda and co-workers. Also shown are the non-covalent interactions suspected to determine the enantioselectivity of the reaction: repulsion between the oxygen of the carbonyl group of the fluoroketone and aryloxy oxygen of the catalyst complex in red; attraction between the carbonyl oxygen and the ammonium proton of the organoboron reagent in purple; attraction between a fluorine of the fluoroketone and the ammonium proton in green; repulsion of a fluorine of the fluoroketone and the aryloxy oxygen of the catalyst complex in blue. 42

14 Interaction energies are sorted by the strength of the counterpoise-corrected MP2/aTZ estimate within each reaction family (A, B, and C) and are plotted relative to the lowest interaction energy estimated within each model chemistry (exchange-scaled sSAPT0/jaDZ and MP2/aTZ). 43

15 Exchange-scaled A-SAPT/jaDZ analysis of the interactions in transition states for Reaction A in terms of electrostatics, exchange-repulsion, induction, and dispersion. Each panel corresponds to an atom-atom interaction depicted in Figure 13. 47

16 Exchange-scaled A-SAPT/jaDZ analysis of the interactions in transition states for Reaction B in terms of electrostatics, exchange-repulsion, induction, and dispersion. Each panel corresponds to an atom-atom interaction depicted in Figure 13. 48

17 Exchange-scaled A-SAPT/jaDZ analysis of the interactions in transition states for Reaction C in terms of electrostatics, exchange-repulsion, induction, and dispersion. Each panel corresponds to an atom-atom interaction depicted in Figure 13. 49

18 Exchange-scaled F-SAPT/jaDZ analysis of various interactions between the whole organoboron catalyst, phenyl, CNBO "chair", and the $(Me)_2$N-CO-CH-*t*-butyl "chain" of the organoboron catalyst with the phenyl of the substrate in transition states for Reaction B in terms of electrostatics, exchange-repulsion, induction, and dispersion. 52

19 The π faces of IIb (left) do not align as favorably for π-π interactions as they align in IVb (right). 53

20 Exchange-scaled F-SAPT/jaDZ analysis of various interactions between the *t*-butyl, phenyl, and CNBO "chair" of the organoboron catalyst with the *ortho*-methyl group of the substrate in transition states for Reaction C in terms of electrostatics, exchange-repulsion, induction, and dispersion. . . . 56

21 The results for interaction energy calculations on the S22 database using the B2PLYP and LB2PLYP with and without counterpoise correcting in the cc-pVDZ (DZ) basis set. Also, the reference values from the S22B revised benchmarks are plotted. The x-axis corresponds to the S22 dimer number, and the y-axis represents the interaction energy (IE) of the dimer in kcal mol^{-1}. 65

22 The results for interaction energy calculations on the S22 database using the B2PLYP and LB2PLYP with and without counterpoise correcting in the cc-pVTZ (TZ) basis set. Also, the reference values from the S22B revised benchmarks are plotted. The x-axis corresponds to the S22 dimer number, and the y-axis represents the interaction energy (IE) of the dimer in kcal mol^{-1}. 66

23 Mean absolute interaction energy differences between B2PLYP-D3 and LB2PLYP-D3 are plotted for S22 and the hydrogen bonding (HB), dispersion dominated (DD), and mixed interaction (MX) subsets of S22 computed in the cc-pVDZ (DZ) and cc-pVTZ (TZ) basis sets. Both results with (CP) and without the counterpoise correction are presented in kcal mol^{-1}. 67

24 Mean absolute errors vs CCSD(T)/CBS reference values for S22 and the hydrogen bonding (HB), dispersion dominated (DD), and mixed interaction (MX) subsets of S22 computed in the cc-pVDZ (DZ) and cc-pVTZ (TZ) basis sets are presented in kcal mol^{-1}. 73

25 The necessary selections, shown by pink squares, to designate fragment C are shown. Only the border atoms, or atoms that connect to fragments not to be included in the selection, are necessary to include when using the input generation tool. 80

26 After the input generation tool is executed, the fragments are colored. The entire selection of fragment C is shown by the pink squares. Fragment A and B are colored by warm and cool colors, respectively. 81

27 Estimated CCSD(T)/CBS, SAPT0/jun-cc-pVDZ, sSAPT0/jun-cc-pVDZ relative interaction energies, in increasing order of the CCSD(T)/CBS values. CCSD(T)/CBS values were estimated as DF-MP2/aug-cc-pV(T,Q)Z + $\delta_{\mathrm{MP2}}^{\mathrm{CCSD(T)}}$/cc-pVDZ. 86

28 The total electrostatic interaction, as predicted by F-SAPT0/jun-cc-pVDZ, between each fragment and the other monomer is displayed for all considered transition states. Each row consists of transition states leading to the (*S*,*R*) *anti*, (*S*,*S*) *syn*, (*R*,*S*) *ent-anti*, and (*R*,*R*) *ent-syn* transition states respectively. The columns consist of the possible conformations (1,2,3, and 4 respectively) of each transition state. The color scale for the images becomes saturated at ± 25 kcal mol^{-1}. 87

29 The total exchange interaction, as predicted by F-SAPT0/jun-cc-pVDZ, of each fragment with the other monomer is displayed for all considered transition states. Each row consists of transition states leading to the (*S*,*R*) *anti*, (*S*,*S*) *syn*, (*R*,*S*) *ent-anti*, and (*R*,*R*) *ent-syn* transition states respectively. The columns consist of the possible conformations (1,2,3, and 4 respectively) of each transition state. The color scale for the images becomes saturated at ± 25 kcal mol^{-1}. 88

30 The total induction term, as predicted by F-SAPT0/jun-cc-pVDZ, for each fragment interacting with the other monomer is displayed for all considered transition states. Each row consists of transition states leading to the (*S*,*R*) *anti*, (*S*,*S*) *syn*, (*R*,*S*) *ent-anti*, and (*R*,*R*) *ent-syn* transition states respectively. The columns consist of the possible conformations (1,2,3, and 4 respectively) of each transition state. The color scale for the images becomes saturated at ± 25 kcal mol^{-1}. 89

31 The total dispersion interaction, as predicted by F-SAPT0/jun-cc-pVDZ, between each functional group with the other monomer is displayed for all considered transition states. Each row consists of transition states leading to the (*S*,*R*) *anti*, (*S*,*S*) *syn*, (*R*,*S*) *ent-anti*, and (*R*,*R*) *ent-syn* transition states respectively. The columns consist of the possible conformations (1,2,3, and 4 respectively) of each transition state. The color scale for the images becomes saturated at ± 25 kcal mol^{-1}. 90

32 The total interaction, as predicted by F-SAPT0/jun-cc-pVDZ, of each functional group with the other monomer for all considered transition states. Each row consists of transition states leading to the (*S*,*R*) *anti*, (*S*,*S*) *syn*, (*R*,*S*) *ent-anti*, and (*R*,*R*) *ent-syn* transition states respectively. The columns consist of the possible conformations (1,2,3, and 4 respectively) of each transition state. 91

33 The classical Coulombic interaction between the partial charges on each of the atoms involved in the interaction between NCH_2 of the pyrrolidinyl fragment and the CHO of the aldehyde of fragment in kcal mol^{-1} for the (*S*,*R*) *anti* transition states. Here, H1 designates the proton of NCH_2 that is closer to the aldehyde, and H2 designates the proton that is farther from the aldehyde. 92

34 The classical Coulombic interaction between the partial charges on each of the atoms involved in the interaction between NCH_2 of the pyrrolidinyl fragment and the CHO of the aldehyde of fragment in kcal mol^{-1} for the (*S*,*S*) *syn* transition states. Here, H1 designates the proton of NCH_2 that is closer to the aldehyde, and H2 designates the proton that is farther from the aldehyde. 93

35 The classical Coulombic interaction between the partial charges on each of the atoms involved in the interaction between NCH_2 of the pyrrolidinyl fragment and the CHO of the aldehyde of fragment in kcal mol^{-1} for the (*R*,*S*) *ent-anti* transition states. Here, H1 designates the proton of NCH_2 that is closer to the aldehyde, and H2 designates the proton that is farther from the aldehyde. 94

36 The classical Coulombic interaction between the partial charges on each of the atoms involved in the interaction between NCH_2 of the pyrrolidinyl fragment and the CHO of the aldehyde of fragment in kcal mol^{-1} for the (*R*,*R*) *ent-syn* transition states. Here, H1 designates the proton of NCH_2 that is closer to the aldehyde, and H2 designates the proton that is farther from the aldehyde. 95

37 Deformation energies estimated at the CCSD(T)/CBS level are plotted against $(R-R_e)^2$, where R is the O–H bond length in the carboxylic acid group, and R_e is its value in the isolated enamine intermediate. CCSD(T)/CBS values were estimated as DF-MP2/aug-cc-pV(T,Q)Z + $\delta_{\mathrm{MP2}}^{\mathrm{CCSD(T)}}$/cc-pVDZ. 96

38 The chosen reference method, counterpoise-corrected MP2/aTZ was tested against counterpoise-corrected ωB97X-V/aTZ on two systems where π-π interactions were suspected, due to aligned π faces. MP2/aTZ estimates are within 1 kcal mol^{-1} of ωB97X-V/aTZ indicating no large overbinding. . . 97

39 Interaction energies are plotted relative to the most negative interaction energy estimated within each model chemistry (SAPT0/jaDZ, exchange-scaled sSAPT0/jaDZ, counterpoise-corrected MP2/aTZ, counterpoise-corrected MP2/aTZ + counterpoise-corrected δ[CCSD(T)/DZ]). 98

40 Interaction energies are plotted relative to the most negative interaction energy estimated within each model chemistry (SAPT0/jaDZ, exchange-scaled sSAPT0/jaDZ and MP2/aTZ). 99

41 The relative deformation energy of each transition state, as computed with MP2/aTZ, is plotted with respect to the minimum deformation energy in each reaction family (A, B, and C). 100

SUMMARY

This thesis focuses on pioneering the application of symmetry-adapted perturbation theory (SAPT) to organocatalysis. Understanding the effect of non-covalent interactions in transition states would enable rational catalyst design. Industrial reactions that produce chiral products would be benefited by greater selectivity toward the desired enantiomer or stereoisomer. One way to achieve this is by optimizing the non-covalent interactions in the transition state such that the transition states leading to the preferred products are the most stabilized. SAPT allows for the analysis of the fundamental components of non-covalent interactions between two monomers in the transition state: electrostatics, exchange-repulsion, induction/polarization, and London dispersion. Its variants, which are also applied throughout, allow for an even more fine-grained analysis of these interactions. Functional group SAPT (F-SAPT) and atomic SAPT (ASAPT) provide the aforementioned decomposition of the interaction energy into contributions of pairs of functional groups (F-SAPT) or pairs of atoms (ASAPT).

For the first time, we have applied F-SAPT and ASAPT to organocatalysis in order to quantify non-covalent interactions in transition states of two reactions. The first reaction, in which we applied only F-SAPT, was the proline-catalyzed intermolecular aldol reaction between benzaldehyde and cyclohexanone, according to the Houk-List mechanism.[3] In this reaction, a favorable electrostatic interaction between the $NCH^{\delta+}$ group of the enamine intermediate and the $^{\delta-}O{=}C$ of benzaldehyde was thought to be of central importance in determining the dominant product of the reaction. However, direct quantification of this contact through the use of F-SAPT revealed that this contact is destabilizing in all transition states, and further analysis revealed that large negative charges on the close nitrogen and oxygen atoms was responsible for this destabilization. The quantification of a C–H/π interaction between the cyclohexene group of the enamine intermediate and the

benzaldehyde phenyl ring confirmed the presence of this interaction in the form of stabilizing dispersion and electrostatics; however, unfavorable exchange-repulsion causes this contact to be destabilizing overall. A contact between an *ortho*-hydrogen of the phenyl and an oxygen of the carboxylic acid group of the enamine was confirmed to be stabilizing by F-SAPT. Ultimately, F-SAPT analysis showed that non-covalent stabilization was not the primary factor in determining the dominant product of the reaction.

Additionally, F-SAPT and A-SAPT were applied to an enantioselective allyl addition to a fluoroketone. Experimentalists hypothesized that certain non-covalent interactions between specific atoms and functional groups in the transition states contributed to the enantioselectivity of the reaction. Our A-SAPT analysis confirmed the presence of the favorable H $\cdots$ F contact that was thought to be most influential in determining the experimentally observed product ratio. The presence of other atom-atom contacts used to explain the relative product ratios were also confirmed on the basis of A-SAPT analysis. F-SAPT was used to show that interactions between the substrate phenyl or *ortho*-methyl phenyl groups and phenyl of the organoboron catalyst, which were previously thought to be unfavorable due to steric hindrance, are actually favorable π-π interactions.

These applications of F-SAPT have also provided inspiration for the creation of tools to automate input generation. The job set up for these types of analyses can be quite a tedious process for larger transition states in organocatalysis. Software has been written to allow users to make custom inputs through the use of a popular molecule viewer's graphical user interface. In addition to this software development, the design and substitution of a low-cost method, local second-order Møller-Plesset perturbation theory (LMP2), for a more costly method, second-order Møller-Plesset perturbation theory (MP2), in double hybrid density functional theory is discussed. A preliminary analysis of its accuracy on a small database of non-covalent interactions suggests that interaction energies of the canonical method may be routinely reproduced.

CHAPTER I

INTRODUCTION

This chapter will introduce the theory necessary to understand the rest of the thesis. For further clarification, it is recommended that the reader consult external resources, such as "Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory" by Attila Szabo and Neil Ostlund.[72]

1.1 Brief Introduction to Electronic Structure Theory

1.1.1 Schrödinger Equation

Electronic structure theory seeks to describe molecular interactions at the subatomic level through consideration of the distribution of electrons with respect to stationary nuclei, which requires the application of quantum mechanics. For many electronic structure theory applications, and this thesis, the non-relativistic time-independent Schrödinger equation is used.

$$\hat{H}\Psi = E\Psi \tag{1}$$

In the equation above, the operator $\hat{H}$ is the non-relativistic time-independent Hamiltonian, and Ψ and E are the wavefunction and energy of the system, respectively.

$$\hat{H} = -\frac{1}{2}\sum_i \nabla_i^2 - \sum_A \frac{1}{2M_A}\nabla_A^2 - \sum_{A,i}\frac{Z_A}{r_{Ai}} + \sum_{A>B}\frac{Z_A Z_B}{R_{AB}} + \sum_{i>j}\frac{1}{r_{ij}} \tag{2}$$

The Hamiltonian defined above accounts for all non-relativistic energy contributions in a molecular system. $\frac{1}{2}\sum_i \nabla_i^2$ represents the sum of the kinetic energy of all electrons i, where ∇^2 is the Laplacian of the i^{th} electron. $\sum_A \frac{1}{2M_A}\nabla_A^2$ represents the sum of the kinetic energy of each nucleus A, and M_A is the mass of each nucleus. $\sum_{A,i}\frac{Z_A}{r_{Ai}}$ is the Coulombic attraction of each nucleus A with each electron i. $\sum_{A>B}\frac{Z_A Z_B}{R_{AB}}$ is the Coulombic repulsion between all

pairs of nuclei A and B, otherwise known as the nuclear repulsion energy. Finally, $\sum_{i>j} \frac{1}{r_{ij}}$, similarly, is the Coulombic repulsion between all pairs of electrons i and j.

In typical applications of this equation, and in the following work, the Born-Oppenheimer Approximation is invoked. This approximation assumes that the nuclei of a system remain in a stationary position R relative to the electronic coordinates r. This is a sensible approximation due to electrons being orders of magnitude less massive than atomic nuclei (mass of proton or neutron: 1.67 x 10^{-27} kg, mass of electron: 9.11 x 10^{-31} kg). The Born-Oppenheimer Approximation implies that the kinetic energy of the nuclei, $\sum_A \frac{1}{2M_A} \nabla_A^2$, is zero since a stationary position is assumed for the nuclei with respect to the electrons. The Born-Oppenheimer Approximation also implies that the nuclear repulsion energy piece, $\sum_{A>B} \frac{Z_A Z_B}{R_{AB}}$, will be constant since R_{AB} is fixed. Only the "electronic Hamiltonian" will be considered in the subsequent discussion:

$$\hat{H} = -\frac{1}{2} \sum_i \nabla_i^2 - \sum_{A,i} \frac{Z_A}{r_{Ai}} + \sum_{i>j} \frac{1}{r_{ij}}. \tag{3}$$

With the electronic Hamiltonian being given, the main task of electronic structure theory computations is to solve for the rest of the Schrödinger equation, the wavefunction Ψ and the energy E. Ψ represents an eigenstate, and E represents an eigenvalue, or energy of the molecular system. With the wavefunction, useful properties of a molecular system can be derived, such as multipole moments or polarizabilities, and the energy of the system will be obtained. The following sections are concerned with techniques for finding the wavefunctions and the associated energies.

1.1.2 Hartree-Fock Theory

Hartree-Fock is the most approximate *ab initio* method used to solve the electronic Schrödinger Equation. Through this method, one can arrive at an approximate wavefunction, or rather, molecular orbitals. This approximation is equivalent for solving for the energy of a molecular system in a mean field of the electrons.

The electronic Hamiltonian described above can be written in two parts:

$$h(i) = -\frac{1}{2}\nabla_i^2 - \sum_A \frac{Z_A}{r_{iA}} \tag{4}$$

$$v(i,j) = \frac{1}{r_{ij}} \tag{5}$$

The first operator $h(i)$ is known as the one-electron operator because it accounts for the energetics of a single electron. As discussed in the previous section, this operator includes the kinetic energy of the i^{th} electron and the interaction of the i^{th} electron with each nucleus, A. The second operator is a two-electron term which accounts for the interaction of electron i with electron j. These two operators in addition to the nuclear repulsion energy make up the Hamiltonian.

Hartree-Fock works under the assumption that the wavefunction, Ψ, can be represented by single Slater determinant, shown below. This is a specially crafted mathematical construct that enforces the antisymmetry of the wavefunction, in accordance with Pauli exclusion principle for fermions.

$$\Psi = \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_1(x_1) & \chi_2(x_1) & \dots & \chi_N(x_1) \\ \chi_1(x_2) & \chi_2(x_2) & \dots & \chi_N(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_1(x_N) & \chi_2(x_N) & \dots & \chi_N(x_N) \end{vmatrix} \tag{6}$$

When a Slater determinant is used in combination with the electronic Hamiltonian, the energy of the system can be written in terms of sums over occupied orbitals a and b like:

$$\langle \Psi | \hat{H} | \Psi \rangle = \sum_a \langle a|h|a\rangle + \frac{1}{2}\sum_{ab} \langle ab||ab\rangle \tag{7}$$

Clemens Roothan developed a method of finding the orbitals, or Slater determinant, that gives the minimum energy, which is known to be the best guess due to the variational principle.

$$FC = SC\epsilon \tag{8}$$

Here, F is what is known as the Fock operator, C is the matrix of orbital coefficients, S is the overlap matrix, and ϵ is the energy of the system.

$$S_{\mu\nu} = (\mu|\nu) \tag{9}$$

The above equation is known as a psuedo-eigenvalue equation because the Fock operator, F, depends on the eigenstate, C.

$$F_{\mu\nu} = h_{\mu\nu} + \sum_{i} \sum_{\lambda\sigma} C^{\star}_{\lambda i} C_{\sigma i} (2[\mu\nu|\lambda\sigma] - [\mu\sigma|\lambda\nu]) \tag{10}$$

This equation can be solved for iteratively. Iterations will terminate once a suitable threshold is reached, i.e. a small numerical change in energy or orbitals.This Self Consistent Field procedure, similarly to other methods of computational chemistry, provides the lowest energy possible by fitting the orbital coefficients C to a set of given basis functions $\tilde{\chi}$.

$$\chi_i = \sum_{\mu} C_{\mu i} \tilde{\chi}_{\mu} \tag{11}$$

For this thesis, only Gaussian type orbitals (GTO) are used for basis sets.

$$\tilde{\chi}^{GTO}_{\mu}(x, y, z) = N x^l y^m z^n e^{-\alpha r^2} \tag{12}$$

Here, x, y, and z represent coordinates in space, $l + m + n$ sum up to the total angular momentum of the orbital represented, N is a normalization factor, and r is the Euclidean distance from the atomic center. In general, the more flexibility afforded by the basis set, i.e. more functions, the more accurate answer can be achieved.

1.1.3 Symmetry-Adapted Perturbation Theory

Symmetry-adapted perturbation theory (SAPT) will be used throughout this thesis.[73] Namely, zeroth-order SAPT (SAPT0) is used in this work. This approximately equates to using a Hartree-Fock wavefunction to describe both monomers and using second-order perturbation theory to describe the interaction between the monomers. SAPT treats the

interaction between two monomers as a perturbation and solves for the corresponding interaction energy. The primary benefit of using SAPT is the energy decomposition it provides in terms of electrostatics, exchange-repulsion, induction/polarization, and London dispersion. This makes it a useful tool for probing the physical origins for interactions throughout chemistry.

In addition to SAPT, two partitions of this method are used here: atomic SAPT (ASAPT) and functional-group SAPT (F-SAPT). Essentially, these two methods offer a finer grained analysis of the interactions between monomers. A standard SAPT procedure, with exactly two monomers, is still performed, and the canonical overall interaction energy will be recovered by both methods. ASAPT yields the decomposed interaction energy between a pair of atoms, with each atom of the pair belonging to a separate monomer in the underlying SAPT computation. F-SAPT yields the decomposed interaction energy between pairs of functional groups, with each functional group of the pair belonging to a separate monomer of the underlying SAPT computation.

The F-SAPT interpretation ensures user-designated functional groups retain an integral number of localized electrons and protons and treats the σ bond linking these functional groups (link bond) as retaining a proton and electron from each of the atoms involved in the bond. Adding a proton from each atom involved is necessary to make the link bond fragments neutral and avoid large oscillations in the partitioning of SAPT electrostatics. To avoid treating the link bond as a separate entity in F-SAPT analysis, it is fractionally assigned to the functional groups it bounds in what is known as "reduced analysis". The assignment can be weighted by the ratio of orbital atomic charges of the atoms that the fragmented bond is between (links by charge), or half of the link bond can be assigned to each fragment involved in the bond (links 50-50). The results for either partitioning are qualitatively essentially the same when chemically separable fragments are chosen.[50] Here, "chemically separable" refers to functional groups that are linked by at most a σ bond. This precludes functional groups that are bound by aromatic, double, *etc.* bonds because the multiple bond provides resonance between the functional groups, so that they are not electronically distinct and do not necessarily each contain an integer number of electrons as

assumed by F-SAPT. For further explanations, the reader should reference work by Parrish *et al.*[50,51]

The SAPT Hamiltonian can be written like:

$$\hat{H} = \hat{F} + \hat{W} + \hat{V} \tag{13}$$

Here, $\hat{F}$ represents the sum of the Fock operators for each monomer, $\hat{W}$ represents the sum of intramonomer correlation operators for each monomer, and $\hat{V}$ represent the interaction between the monomers. For the SAPT0 method, which is the only version of SAPT utilized by this thesis, the interaction energy computed can be represented as follows:

$$E_{int}^{SAPT0} = E_{int}^{HF} + E_{disp}^{(20)} + E_{exch-disp}^{(20)}. \tag{14}$$

The E_{int}^{HF} term is the Hartree-Fock interaction energy and is computed in a SAPT computation using the supersystem approach:

$$E_{int}^{HF} = E_{dimer}^{HF} - E_{monomerA}^{HF} - E_{monomerB}^{HF}. \tag{15}$$

E_{dimer}^{HF} is the energy of both monomers in the dimer geometry, the "supersystem", and $E_{monomerA}^{HF}$ and $E_{monomerB}^{HF}$ are the energies of each monomer in isolation. The last two terms of the sum above ($E_{disp}^{(20)}$ and $E_{exch-disp}^{(20)}$) represent the London dispersion contribution of the aforementioned energy decomposition. The E_{int}^{HF} term can be decomposed into the additional terms (electrostatics, exchange-repulsion, and induction/polarization) like:

$$E_{int}^{HF} = E_{elst}^{(10)} + E_{exch}^{(10)} + E_{ind,resp}^{(20)} + E_{exch-ind}^{(20)} + \delta E_{int}^{HF} \tag{16}$$

The first term in the above equation represents the contribution from electrostatic interactions, the second term represents the contribution from exchange-repulsion, and the last three terms summed together represent the contributions from induction/polarization. The δE_{int}^{HF} term is purely the contribution needed to retrieve the Hartree-Fock interaction energy and is included in the induction/polarization component throughout this thesis. Electrostatic contributions can be stabilizing (charge-penetration, favorable Coulombic

interaction, *etc.*) or destabilizing (unfavorable Coulombic interaction or multipole interaction, *etc.*). Exchange-repulsion contributions should always be destabilizing, and induction/polarization and London dispersion contributions should always be stabilizing. However, it must be noted that for ASAPT and F-SAPT, some individual contributions to the induction/polarization term for a given pair of atoms or functional groups can actually be destabilizing, but the sum of these contributions will result in an overall stabilizing induction/polarization term.[51]

1.2 Post Hartree-Fock Methods of Electron Correlation

Electron correlation is formally defined as the energy difference of the exact non-relativistic electronic energy and Hartree-Fock energy (E_{HF}) of a system.

$$E_{correlation} = E_{exact} - E_{HF} \tag{17}$$

In order to capture this energy, more advanced theories are required.

1.2.1 Møller-Plesset Perturbation Theory

Møller-Plesset Perturbation Theory (MPPT)[46] is the cheapest *ab initio* method that may be used to account for electron correlation. It uses perturbation theory, typically to second-order (known as MP2), to account for electron correlation as a perturbation $\lambda\hat{V}$ to the aforementioned Hartree-Fock Hamiltonian $\hat{H}_0$, where λ is an arbitrary parameter to control the size of the perturbation $\hat{V}$.

$$\hat{H} = \hat{H}_0 + \lambda\hat{V} \tag{18}$$

The perturbation is defined as the difference in the electronic Hamiltonian and the Hartree-Fock potential.

$$\hat{V} = \sum_{i<j} \frac{1}{r_{ij}} - \hat{V}^{HF} \tag{19}$$

The wavefunction and energy are represented as a power series expansion in λ.

$$(\hat{H}_0 + \lambda\hat{V})(\sum_{i=0}^{m} \lambda^i |\Psi_n^{(i)}\rangle) = (\sum_{i=0}^{m} \lambda^i E_n^{(i)})(\sum_{i=0}^{m} \lambda^i |\Psi_n^{(i)}\rangle) \tag{20}$$

The equation to be solved can be formed by equating all terms with λ to the same power k, where k is considered the order of the perturbation theory. As stated above, this is typically truncated at $k = 2$, or second-order perturbation theory. For instance, when $k = 2$, the following perturbation expansions is formed:

$$H_0|\Psi_n^{(2)}\rangle + \hat{V}|\Psi_n^{(1)}\rangle = E_n^{(0)}|\Psi_n^{(2)}\rangle + E_n^{(1)}|\Psi_n^{(1)}\rangle + E_n^{(2)}|\Psi_n^{(0)}\rangle \tag{21}$$

Solving this equation for the second-order energy, $E^{(2)}$, yields the following:

$$E_n^2 = \langle\Psi_n^{(0)}|\hat{V}|\Psi_n^{(1)}\rangle. \tag{22}$$

Now, only $\Psi_n^{(1)}$ needs to be defined to determine the MP2 correlation energy, which can be solved for using the first order expansion.

$$|\Psi_n^{(1)}\rangle = \sum_{m\neq n} \frac{\langle\Psi_m^{(0)}|\hat{V}|\Psi_n^{(0)}\rangle}{E_n^{(0)} - E_m^{(0)}}|\Psi_m^{(0)}\rangle \tag{23}$$

By combining these definitions, the MP2 energy can be expressed.

$$E_0^{(2)} = \sum_{m\neq 0} \frac{|\langle\Psi_0^{(0)}|\hat{V}|\Psi_m^{(0)}\rangle|^2}{E_0^{(0)} - E_m^{(0)}} \tag{24}$$

The above equation can be expressed in terms of two-electron integrals in the molecular orbitals. These are computed *via* a transformation of the atomic orbital integrals computed by Hartree-Fock. More on this operation is discussed in the section pertaining the density fitting approximation, 1.3. Typically, higher than second-order perturbation theory is avoided due to the fact that it is not systematically improvable nor necessarily convergent at higher levels. Even second-order perturbation is subject to divergence problems if $E_0^{(0)}$ and $E_m^{(0)}$ become close in energy in the above expression. However, MP2 generally serves as a good estimate of electron correlation effects despite being notorious for overbinding π-π interactions.[1]

1.2.2 Local MP2

Another way of decreasing the cost of electronic structure computations is through making local correlation approximations. These approximations rely on the idea that electron motions are correlated only if they are in close proximity. The following discussion of local correlation will concern the work on local MP2 of Werner and Pulay.[54,62,81]

The general algorithm of local MP2 is performed in the following steps:

1. Localize Orbitals
2. Determine Pair Domains and Classes
3. Compute Two-Electron Integrals
4. Solve for Amplitudes Iteratively

Localizing the orbitals involves transforming the canonical orbitals using a unitary transformation, which implies the energy of the system does not change, in order to make them more amenable to chemical intuition and less diffuse. By making the orbitals localized, meaningful estimates can be made to predict the strength of orbital interactions *a priori*. This procedure can be done through multiple well known localization algorithms, for example those of Boys[10] or Pipek-Mezey.[53] Local molecular orbitals (LMOs) are defined in the atomic orbital basis $\{\chi_\mu\}$ by a matrix of coefficients L:

$$|\phi_i\rangle = \sum_\mu |\chi_\mu\rangle L_{\mu i}, \tag{25}$$

where $|\phi_i\rangle$ is now being used to represent an LMO. From these local orbitals, virtual orbitals are produced in the form of projected atomic orbitals (PAOs) that are non-orthogonal functions that span the virtual space:

$$|\tilde{\phi}_r\rangle = (1 - \sum_i^m |\phi_i\rangle\langle\phi_i|)|\chi_\mu\rangle = \sum_\mu |\chi_\mu\rangle P_{\mu r}. \tag{26}$$

A tilde will be used to signify terms that are in the PAO basis. Additionally, the matrix P can be computed through the following matrix equation:

$$P = 1 - LL^{\dagger}S, \tag{27}$$

where S is the standard overlap matrix. With the LMOs and PAOs defined, it is now possible to form orbital domains. Orbital domains make it possible to limit the number of atomic orbital integrals that are transformed into the LMO basis. These orbital domains, $[i]$, are a subset of PAOs that are the most relevant for each LMO. They are formed by first sorting all atoms by their Mulliken charge contribution to LMO i. Then for each LMO i, atoms from the sorted list are added one at a time to a list of relevant atoms until the Boughton-Pulay threshold (BP Threshold) is met.[9]

$$1 - \int (\phi_i - \tilde{\phi}_i)^2 dr \geq \text{BP Threshold} \tag{28}$$

Here, $\tilde{\phi}_i$ is an approximation of the orbital represented by the selected atoms.

$$|\tilde{\phi}_i\rangle = \sum_{A \in [i]} \sum_{\mu \in A} |\chi_\mu\rangle \tilde{L}_{\mu i} \tag{29}$$

Once the BP Threshold is met, the PAOs, r or s, that are formed by atomic orbitals at an atom center in the list of relevant atoms are considered part of orbital domain $[i]$. Then, united pair domains $[ij]$ are formed by taking the union, $[i] \cup [j]$, of the corresponding pair domains. Each pair domain is classified by a distance-based hierarchy according to the closest contact between atoms in each respective pair domain. This classification is used to determine the theoretical treatment for the electron correlation between each pair of occupied LMOs i and j, and the computational costs are reduced by requiring less rigorous computations for each distance regime:

- Strong Pairs: R $<$ 2 Bohr: Local Coupled Cluster
- Weak Pairs: 2 $\leq$ R $<$ 8 Bohr: Local MP2
- Distant Pairs: 8 $\leq$ R $<$ 15 Bohr: Multipole Approximation to Local MP2
- Very Distant Pairs: R $\geq$ 15 Bohr: Neglect

These approximations, together with the restriction of the virtual orbital space to only the PAO's in the united pair domain $[ij]$ for each pair of occupied LMOs i and j, makes the computational cost independent of molecular size. In this thesis, integrals computed in the "Strong Pair" regime were computed with MP2 instead of coupled cluster. All integral transformations can be density fitted, and the problem size can be further reduced by applying techniques described by Werner, Manby, and Knowles.[81]

The local MP2 energy can be written in terms of the two-electron integrals, $\tilde{K}^{ij}_{rs} = (ri|sj)$, as shown below.

$$E = \sum_{ij \in P} \sum_{r,s \in [ij]} (2\tilde{T}^{ij}_{rs} - \tilde{T}^{ji}_{rs})\tilde{K}^{ij}_{rs} \tag{30}$$

Here, P is the list of ij pairs that considered in the LMP2 computation, and r and s are the PAOs associated with the united pair domain $[ij]$. Local MP2 is an iterative method which aims to update the amplitudes, $\tilde{T}^{ij}_{rs}$, such that the residual term as derived from the Hylleraas functional is minimized.[62]

$$R^{ij} = \tilde{K}^{ij} + \tilde{F}\tilde{T}^{ij}\tilde{S} + \tilde{S}\tilde{T}^{ij}\tilde{F} - \sum_k \tilde{S}[F_{ik}\tilde{T}^{kj} + F_{kj}\tilde{T}^{ik}]\tilde{S} \tag{31}$$

In the above equation, $\tilde{F}$ and $\tilde{S}$ are the Fock and overlap matrix in the PAO basis, respectively. Work by Werner and company suggests that it may be faster to compute the summation term with small matrix multiplies, $\tilde{S} \times \tilde{T}^{kj} \times \tilde{S}$, for each k inside the summation. However, our studies suggest that it is much faster on modern hardware to form the matrix produced by the summation and compute one large matrix multiply outside of the summation.

In addition to the density fitting approximation mentioned above, a multipole expansion can be used to approximate the two-electron integrals $(ri|sj)$ as shown below.

$$\tilde{K}^{ij}_{rs} \approx \sum_{lm} \sum_{l'm'} Q^{ri}_{lm} U^{ir,js}_{lm,l'm'} Q^{sj}_{l'm'} \tag{32}$$

Q^{ri}_{lm} is the m^{th} multipole moment of order l computed at the center of the charge distribution $\rho_{ir} = \phi_i\tilde{\phi}_r$ and $U^{ir,js}_{lm,l'm'}$ is the interaction operator. Typically, this expansion

is truncated at $l = 3$.[28] This interaction operator computes the interaction between the multipoles Q_{lm}^{ri} and $Q_{l'm'}^{sj}$.

$$U_{lm,l'm'}^{ir,js} = \frac{V_{lm,l'm'}^{ir,js}}{R^{l+l'+1}} \tag{33}$$

R is the distance between the centers of the interacting multipole, and l and l' are the orders of the multipole expansion about the center of charge distribution ρ_{ri} and ρ_{sj}, respectively. $V_{lm,l'm'}^{ir,js}$ describes the interaction between the multipole centers and its forms can be found in "The Theory of Intermolecular Forces" by Anthony Stone.[71] Also, "asymmetric pair domains" are required for this expansion to converge. "Asymmetric pair domains" are the subsets of PAOs (r and s) in the pair domain $[ij]$ that only belong to exactly one of the orbital domains $[i]$ or $[j]$.[28] This is due to the inverse dependence on R in the interaction operator, $R^{l+l'+1}$. Charge distributions that are too close will cause the interaction to blow up.

For integrals approximated with a multipole expansion, the multipole expansion can be computed at some point of origin, $\mathbf{0}$, and inexpensively translated to other centers $\mathbf{R}$.[29]

$$Q_{l_x l_y l_z}^{ri:\mathbf{R_{ri}}} = \sum_{k_x k_y k_z} (-\mathbf{R_{rix}})^{l_x - k_y} (-\mathbf{R_{riy}})^{l_y - k_y} (-\mathbf{R_{riz}})^{l_z - k_z} Q_{k_x k_y k_z}^{ri:\mathbf{0}} \tag{34}$$

Here, $Q_{k_x k_y k_z}^{ri:\mathbf{0}}$ is the multipole moment of charge distribution ρ_{ri} determined at the origin and has been converted from spherical polar coordinates to Cartesian coordinates through transformations that can be found in "The Theory of Intermolecular Forces" by Anthony Stone.[71] $Q_{l_x l_y l_z}^{ri:\mathbf{R_{ri}}}$ is the Cartesian multipole of charge distribution ρ_{ri} that has been translated to center $\mathbf{R_{ri}}$. $\mathbf{R_{rix}}$ is the x coordinate of the center, $\mathbf{R_{ri}}$, of the translated multipole. Similarly, $\mathbf{R_{riy}}$ and $\mathbf{R_{riz}}$ are the y and z coordinates. After translation, these Cartesian moments can be converted back into spherical polar coordinates for use in the interaction formula above.

1.2.3 Coupled-Cluster Theory

Coupled-Cluster Theory[55,56,65] is a more rigorous method used to account for electron correlation and is often used to compute what are considered to be the most accurate

energetics in quantum chemistry, known as the "Gold Standard". It features an exponential ansatz that is expanded in a Taylor series and dramatically adds to the number of terms present and the accuracy of the energetic description.

$$|\Psi\rangle = e^{\hat{T}}|\Phi_0\rangle \tag{35}$$

As previously stated, this wavefunction can be represented by Taylor expansion in $\hat{T}$.

$$e^{\hat{T}}|\Phi_0\rangle = (1 + \frac{1}{1!}\hat{T}^1 + \frac{1}{2!}\hat{T}^2 + \frac{1}{3!}\hat{T}^3 + ...)|\Phi_0\rangle \tag{36}$$

$\hat{T}$ is known as the cluster operator and controls the number of excitations of the wavefunction that are taken into account.

$$\hat{T} = \hat{T}_1 + \hat{T}_2 + \hat{T}_3 + ... \tag{37}$$

The order of coupled cluster is determined by where the cluster operator is truncated. For instance, if $\hat{T}$ is terminated at T_2, only up to double excitations ($\hat{T}_1$ and $\hat{T}_2$) are taken into account, and this is known as coupled cluster singles and doubles (CCSD).

$$\hat{T}_1|\Phi_0\rangle = \sum_{i,a} t_i^a|\Phi_i^a\rangle \tag{38}$$

$$\hat{T}_2|\Phi_0\rangle = \sum_{i>j,a>b} t_{ij}^{ab}|\Phi_{ij}^{ab}\rangle \tag{39}$$

Here, t_i^a and t_{ij}^{ab} are what is known as the singles and doubles amplitudes, respectively, which are the coefficients of the CCSD wavefunction. $|\Phi_i^a\rangle$ is a singly excited determinant which is generated by substituting occupied molecular orbital i for a virtual orbital a. Similarly, $|\Phi_{ij}^{ab}\rangle$ is a doubly excited determinant and is generated by making the same substitution of i for a in addition to the substitution of occupied molecular orbital j for a virtual molecular orbital b. This level of coupled-cluster is even more accurate than MP2 and can be systematically improved by increasing the level of excitations taken into account (for example triple or quadruple excitations); however, this will result in a prohibitively expensive

computation on modern hardware in terms of wall time and memory. The aforementioned "gold standard" of computational chemistry is achieved by performing CCSD plus a perturbative triples estimate, known as CCSD(T). This is already often a prohibitively expensive computation and can only be done routinely on systems that are less than 100 atoms, even when a host of approximations are made.

1.3 Density Fitting

A common bottleneck in electronic structure computations is the computation and storage of two-electron integrals. Density fitting is a widely used to technique to greatly reduce the memory requirements of storing these integrals by reducing these integrals from four-index to three-index objects.[27,30,31,63,64,80] Two-electron integrals are represented in the following manner:

$$(pq|rs) = \int d\mathbf{r_1} \int d\mathbf{r_2} \phi_p(\mathbf{r_1}) \phi_q(\mathbf{r_1}) \frac{1}{\mathbf{r_{12}}} \phi_r(\mathbf{r_2}) \phi_s(\mathbf{r_2}). \tag{40}$$

These integrals can alternatively written in terms of densities, ρ_{pq}, and approximated by what is known as a fitting basis set χ:

$$\bar{\rho}_{pq}(\mathbf{r}) = \sum_{P}^{N_{fit}} d_P^{pq} \chi_P(\mathbf{r}) \tag{41}$$

Others have shown that fitting for the coefficients d_P^{pq} can be achieved through the equation below.[15,16]

$$d_P^{pq} = \sum_{Q} (pq|Q)[J^{-1}]_{PQ} \tag{42}$$

In the above equation, the three-index quantity $(pq|P)$ is defined as

$$(pq|P) = \int d\mathbf{r_1} \int d\mathbf{r_2} \phi_p(\mathbf{r_1}) \phi_q(\mathbf{r_1}) \frac{1}{\mathbf{r_{12}}} \chi_P(\mathbf{r_2}), \tag{43}$$

and the J_{PQ} matrix, or Coulomb metric, is defined as

$$J_{PQ} = \int dr_1 \int dr_2 \chi_P(r_1) \frac{1}{r_{12}} \chi_Q(r_2). \tag{44}$$

Now, the original integrals equations can be computed with summation below.

$$(pq|rs) = \sum_{PQ} (pq|P)[J^{-1}]_{PQ}(Q|rs). \tag{45}$$

Performing this approximation formally reduces the scaling of the most expensive two-electron integral transformations in MP2. For instance, the canonical integral transformation of atomic orbitals that is necessary to compute the equations for MP2 is formally $\mathcal{O}(N^5)$.

$$N = N_{occ} N_{AO}^4. \tag{46}$$

However, when density fitting is applied, this scaling is reduced to $\mathcal{O}(N^4)$.

$$N = N_{occ} N_{AO}^2 N_{aux}. \tag{47}$$

Here, N_{occ} is the number of occupied orbitals, N_{virt} is the number of virtual orbitals, N_{aux} is the number of auxiliary functions, and N_{AO} is the number of atomic orbitals. In general, N_{AO} is greater than N_{occ} and N_{virt}, which leads to the speedup seen in practice. Additionally, the storage required for the transformation is reduced from $\mathcal{O}(N^4)$ to $\mathcal{O}(N^3)$. The first transformation step of canonical MP2 requires the costly storage of $(i\nu|\lambda\sigma)$ integrals, a four-index quantity. In comparison, the most memory intensive step of density fitted MP2 is the storage of $(ia|P)$ intermediates, a three-index quantity.

1.4 Organization of Thesis

The rest of this thesis will concern both published and unpublished works that make use of the material in the introduction. The next two chapters will focus on the examination of non-covalent interactions in organocatalysis through the use of SAPT and its variants. After that, the next chapter will focus on the implementation and substitution of local MP2 in a double hybrid DFT, B2PLYP. Finally, algorithms and tools designed to aid and automate input generation for F-ISAPT computations will be discussed.

CHAPTER II

EXAMINATION OF NON-COVALENT INTERACTIONS IN THE HOUK-LIST MECHANISM

2.1 Summary

Rational design of catalysts would be aided by a better understanding of how non-covalent interactions stabilize transition states. Here, we apply the newly-developed Functional-Group Symmetry-Adapted Perturbation Theory (F-SAPT) to quantify non-covalent interactions in transition states of the proline-catalyzed intermolecular aldol reaction between benzaldehyde and cyclohexanone, according to the Houk-List mechanism [Bahmanyar *et al., J. Am. Chem. Soc.* 2003, **125**, 2475]. A recent re-examination of this organocatalytic reaction by Rzepa and co-workers [Armstrong *et al., Chem. Sci.* 2014, **5**, 2057] used electron density analysis to identify three key non-covalent interactions thought to influence stereoselectivity: (1) a favorable electrostatic interaction (originally identified by Houk and List) between the $NCH^{\delta+}$ group of the enamine intermediate and the $^{\delta-}O{=}C$ of benzaldehyde; (2) a C–H/π interaction between the cyclohexene group of the enamine intermediate and the benzaldehyde phenyl ring; (3) a stabilizing contact between an *ortho*-hydrogen of the phenyl and an oxygen of the carboxylic acid group of the enamine. These three interactions have been directly computed using F-SAPT, which confirms the stabilizing interaction between an *ortho*-hydrogen and the carboxylic acid in the (*S*,*S*) and (*R*, *S*) transition state stereoisomers. F-SAPT analysis also finds stabilizing dispersion and electrostatic interactions due to a C–H/π interaction between the cyclohexene and phenyl groups in the (*S*,*S*) and (*R*,*R*) transition states. However, unfavorable exchange-repulsion cancels the attractive terms that favor these stereoisomers. Surprisingly, the interaction thought to be most important for stereoselectivity, the $NCH^{\delta+} \cdots {}^{\delta-}O{=}C$ interaction, is actually found to be repulsive due to the negative charge on the nitrogen. Hence, our results indicate that geometric analysis and/or density-based analysis does not necessarily produce

a reliable picture of non-covalent stabilization. As confirmed by high-level coupled-cluster computations, intermolecular interaction energies are strongest for the (*R*,*R*) transition states, which are not the experimentally favored products. This suggests that at least for this reaction, stereoselectivity is also strongly dependent on the energy required to distort the reacting molecules into the transition state geometry.

2.2 Introduction

Catalytic design would benefit greatly from a more detailed understanding of the factors influencing transition state energetics. As recognized by the "distortion-interaction" model of Ess and Houk[20] and the "activation strain" model of Bickelhaupt and co-workers,[21] the two factors determining the energy of a transition state are (1) the strain energy required to deform the reactants into the transition state geometry, and (2) the stabilization coming from intermolecular interactions between the reactants. Indeed, a seminal paper by Houk, List, and co-workers[3] proposes that non-covalent interactions (NCI) that stabilize transition states can determine the stereoselectivity of proline-catalyzed aldol reactions. Rzepa and co-workers[2] have recently re-examined this reaction and claim that inspection of non-covalent interactions "is shown to be a useful tool for the design of alternative reactants."

In their study, Rzepa and co-workers used the NCIPLOT program,[13] which provides a 3D color-coded map based on the values of the density, the reduced density gradient, and the Laplacian of the density. Johnson et al.[34] found an empirical correlation between various combinations of these quantities and the type and qualitative strength of NCI. Relying on this density-based analysis, Rzepa and co-workers identified additional sources of non-covalent stabilization of transition states in proline-catalyzed aldol reactions.

The density-based NCI analysis provides a simple and intuitive way to visualize intermolecular interactions. However, it does not provide energetics for these interactions. Symmetry-adapted perturbation theory (SAPT)[33,73] does provide energetics for NCI, and it also provides the fundamental components of these interactions: electrostatics, induction/polarization, London dispersion forces, and steric exchange-repulsion. Such information is very helpful in understanding NCI,[70] but for catalytic design, even more fine-grained

information would be useful: one would also like to know the non-covalent stabilization or destabilization *between individual pairs of functional groups*. Fortunately, this information is now available through functional-group SAPT (F-SAPT), recently developed by our group.[50]

Here we apply F-SAPT to the proline-catalyzed aldol reaction in order to quantify the strength of the NCI proposed by Houk and co-workers[3] and by Rzepa and co-workers[2] as being important for transition state stabilization and stereoselectivity for the proline-catalyzed aldol reaction of benzaldehyde and cyclohexanone.

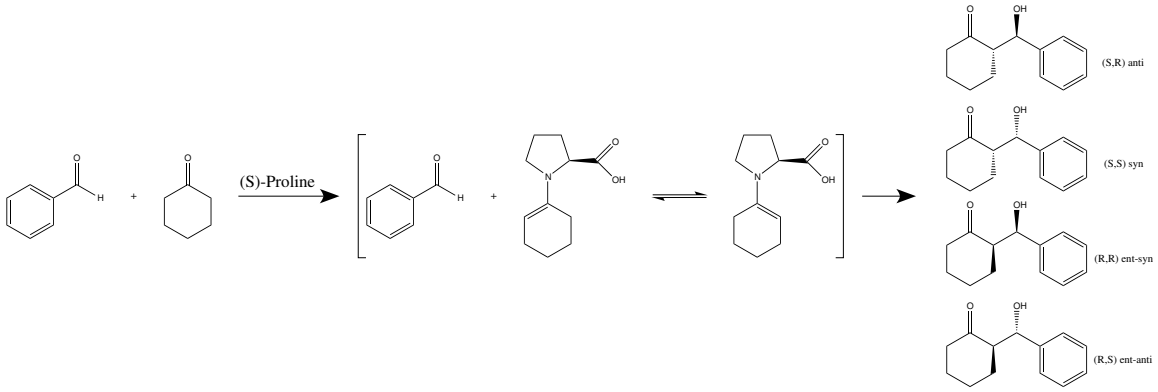


Figure 1: The proline-catalyzed reaction between benzaldehyde and cyclohexanone along with the four possible stereochemical outcomes [(*S*,*R*) *anti*, (*S*,*S*) *syn*, (*R*,*R*) *ent-syn*, and (*R*,*S*) *ent-anti*]. Proline catalyzes the reaction by reacting with cyclohexanone to form a reactive enamine intermediate.

Four different stereoisomers can result from this reaction, as illustrated in Figure 12. The (*S*,*R*) *anti* and (*S*,*S*) *syn* products appear in near equal product ratios, 45-47% and 43-45% respectively.[3] Likewise, the (*R*,*R*) *ent-syn* and (*R*,*S*) *ent-anti* stereoisomers have product ratios of 5-7% and 3-5% respectively.[3] Four stereoisomers are possible because the carbon-carbon double bond that takes part in the C–C bond formation can exist on the same side or the opposite side of the carboxylic acid group in the enamine intermediate, and because the phenyl group of benzaldehyde can be oriented in two different ways relative to the enamine intermediate, as illustrated in Figure 2. Rzepa and co-workers[2] have identified four low-lying conformers for each of these four families of transition states (see Figure 3), leading to a total of 16 relevant possible transition state structures.

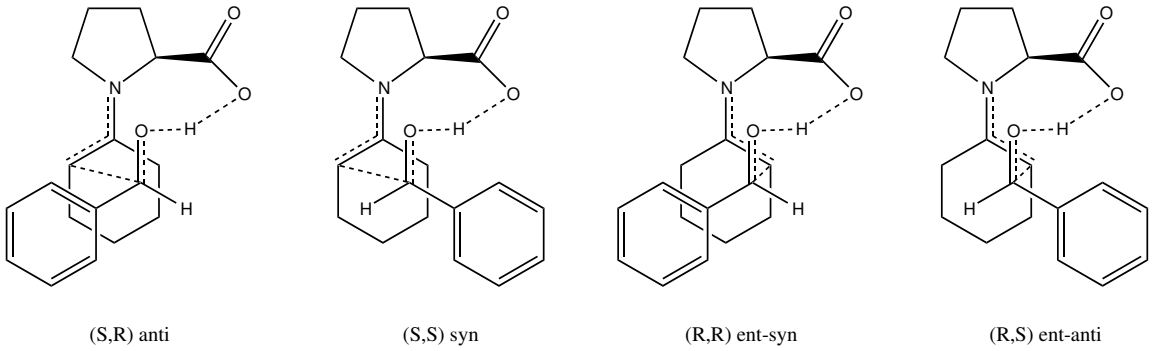


Figure 2: Transition states, without specified conformations, leading to each possible stereoisomer [(*S*,*R*) *anti*, (*S*,*S*) *syn*, (*R*,*R*) *ent-syn*, and (*R*,*S*) *ent-anti*] of the product. The stereochemistry is determined by the relative orientations of the phenyl group of benzaldehyde and the forming carbon-carbon bond.

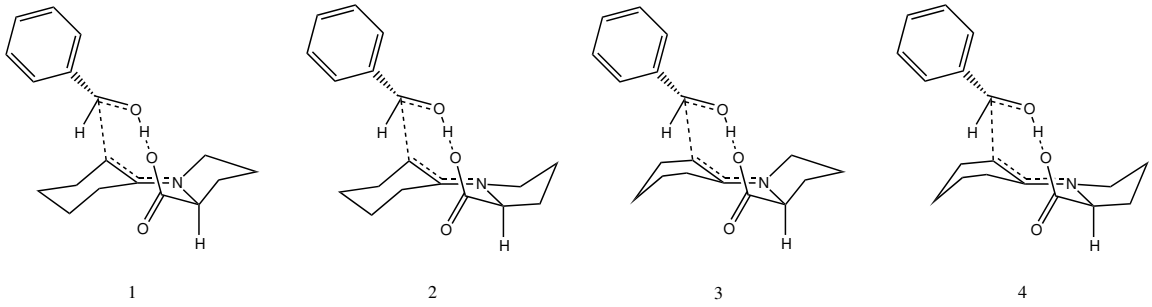


Figure 3: The four low-lying conformations for each transition state stereoisomer. The pyrrolidinyl ring is puckered away from benzaldehyde in conformers 1 and 3, and puckered towards it in conformers 2 and 4. Cyclohexene adopts a chair conformation in conformers 1 and 2, and a boat conformation in conformers 3 and 4.

Here we assess non-covalent stabilization in all 16 of Rzepa's transition state geometries for the proline-catalyzed aldol reaction. We directly compute the strength of interactions postulated[2,3] to be important to transition state stabilization (see Figure 4) including (a) $NCH^{\delta+} \cdots {}^{\delta-}O{=}C$ electrostatic interactions between the NCH group of the pyrrolidinyl fragment and the C=O bond of the aldehyde, (b) C–H/π interactions between the cyclohexene ring of the enamine intermediate and the phenyl group of benzaldehyde, and (c) favorable electrostatic interactions between a ${}^{\delta+}H$ *ortho*-hydrogen of the phenyl group from benzaldehyde and the hydroxyl oxygen of the carboxylic acid group of the enamine intermediate.

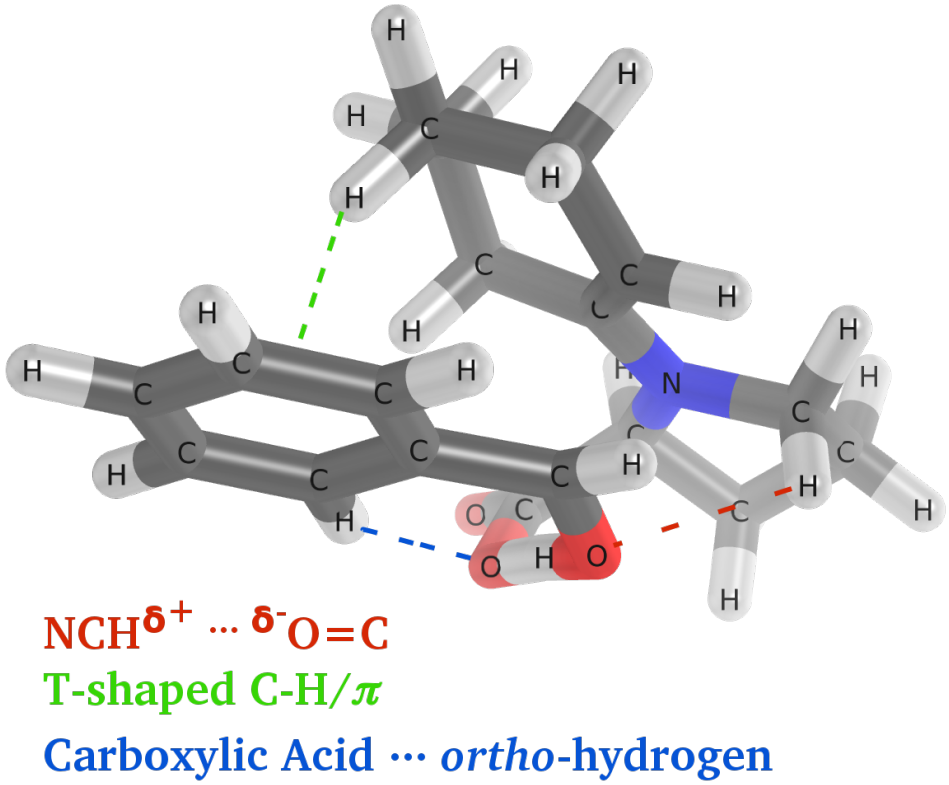


Figure 4: The (*S*,*S*) *syn* 1 transition state, with the three hypothesized sources of non-covalent stabilization: NCH$^{\delta+}$ ··· $^{\delta-}$O=C, C–H/π, and carboxylic acid ··· *ortho*-hydrogen

2.3 Computational Methods

Symmetry-adapted perturbation theory (SAPT)[33] offers the advantage of decomposing molecular interactions into components of electrostatics, exchange-repulsion, induction/polarization, and London dispersion forces. F-SAPT further improves upon this by providing interaction components for each pair of (user-defined) functional groups. Here we use F-SAPT based on the SAPT0 truncation of the perturbation series, which is essentially a second-order intermolecular perturbation theory based on a Hartree–Fock description of the monomers, in a developers' version of Psi4.[74] Despite its lack of intramolecular correlation, SAPT0

provides reliable results for NCI[49] due to cancellation of errors when coupled with a jun-cc-pVDZ basis set,[48] which is the standard Dunning aug-cc-pVDZ basis[17,36] less diffuse functions on hydrogen atoms and d diffuse functions on heavy atoms.

The appropriate fitting basis sets are used for the density fitted algorithms: self-consistent field (SCF) procedure uses jun-cc-pVDZ-JK, and two-body contributions from SAPT0 (dispersion and exchange-dispersion) use jun-cc-pVDZ-RI.[78] Additionally, the core orbitals electrons of heavy atoms were constrained to be doubly occupied ("frozen") in all computations.

The transition state was partitioned into the functional groups shown in Figure 5 for F-SAPT analysis. As seen in Figures 2 and 5, a proton is being transferred from the carboxylic acid of the enamine intermediate to the benzaldehyde reactant. Standard SAPT (and F-SAPT) analysis requires two interacting monomers, and hence we partition the system into the original reactants for the purpose of the F-SAPT procedure. This choice avoids the positive and negative charges that would result from working with the monomers where the proton is accounted as already transferred, and these charges would tend to overwhelm other SAPT terms and thus complicate the analysis. For the F-SAPT analysis, benzaldehyde was considered as a phenyl group plus an aldehyde group, and the enamine intermediate was considered as a five membered pyrrolidinyl nitrogen containing ring, a carboxylic acid group, and a cyclohexene ring. In this work, we use the transition state geometries previously obtained by Rzepa and co-workers.[2]

All molecular pictures generated using F-SAPT are colored with a red-white-blue color palette according to the interaction each functional group experiences with the included functional groups of the other monomer. Red signifies an attractive interaction, and blue signifies a repulsive interaction. The color spectrum in all images becomes saturated at $\pm$ 25 kcal mol^{-1} unless otherwise stated. Our F-SAPT analysis is supplemented in some cases by consideration of atomic charges, which we obtained using natural population analysis from NBO 5.0 in Q-Chem with the def2-TZVP basis set.[22,68,77]

This is the first application of F-SAPT to transition states, and so it is important to assess the reliability of the results, particularly because the intermolecular interaction will be strong in a transition state featuring bond breaking and/or bond forming between the SAPT

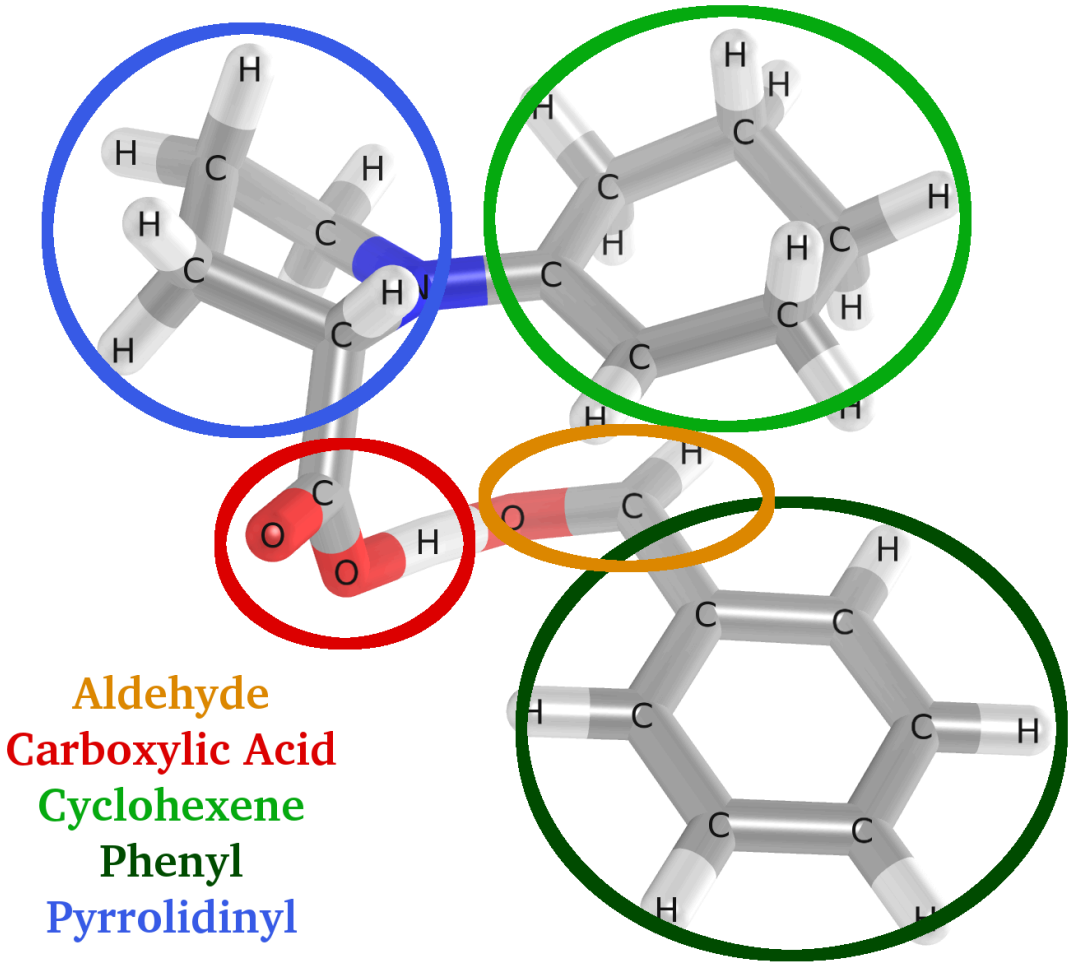


Figure 5: The fragmentation scheme used for all F-SAPT analyses in this work, unless otherwise stated. There are five distinct fragments in total for the two monomers. Enamine intermediate: 1) carboxylic acid, 2) cyclohexene, 3) pyrrolidinyl. Benzaldehyde: 4) aldehyde and 5) phenyl.

monomers, as this one does. Figure 27 of the Ancillary Material compares F-SAPT0/jun-cc-pVDZ interaction energies of the two reacting molecules in the transition state to results from coupled-cluster theory with perturbative triple substitutions [CCSD(T)],[56] estimated in the complete-basis-set (CBS) limit using a focal-point approach.[18] Specifically, we estimate the CBS limit of second-order Møller-Plesset perturbation theory (MP2) and then

correct this for higher-order correlation effects present in CCSD(T) but absent in MP2 by adding the difference between CCSD(T) and MP2, $\delta_{\mathrm{MP2}}^{\mathrm{CCSD(T)}}$, evaluated in a cc-pVDZ basis set. The MP2/CBS values were estimated using a standard two-point Helgaker extrapolation[25] of the correlation energies. MP2 and CCSD(T) results employed density fitting approximations, and CCSD(T) results also employed MP2 frozen natural orbitals,[14] and truncated orbitals with occupation numbers less than 10^{-6}. All of these computations were performed using Psi4.[74]

The comparison in Figure 27 shows that both model chemistries predict nearly identical *trends* with respect to the relative interaction energies; however, F-SAPT0/jun-cc-pVDZ substantially overestimates the overall relative interaction energies (by $\sim$ 5-12 kcal mol^{-1}, or 40-60%). This large discrepancy can be attributed to the large energetic perturbations caused by the transfer of a proton and the formation of a carbon-carbon bond between the two SAPT monomers in the transition state. Fortunately, the energetic ordering of the transition states in terms of their intermolecular interaction energies is in nearly perfect agreement between F-SAPT0/jun-cc-pVDZ and CCSD(T)/CBS, which gives reason to believe F-SAPT is able predict the correct trends of interactions between functional groups. Additionally, we expect the largest errors in the F-SAPT analysis to be for the aldehyde $\cdots$ carboxylic acid pair, where a proton is being transferred, and for the aldehyde $\cdots$ cyclohexene pair, where a C–C bond is being formed. Fortunately, the three hypothesized sources of NCI stabilization in the literature that are the focus of the present study are not between those pairs of functional groups.

2.4 Results and Discussion

Rzepa and co-workers recognized[2] three sources of non-covalent stabilization using a density-based analysis.[34] As illustrated in in Figure 4, they were as follows:

1. Significant stabilization between the oxygen of the aldehyde and a nearby hydrogen on the pyrrolidinyl fragment in (*S*,*R*) *anti* and (*S*,*S*) *syn* transition states [NCH$^{\delta+}$ $\cdots$ $^{\delta-}$O=C]. Houk and co-workers noticed these close contacts in their original study[3] and argued that they were key in the selectivity to the (*S*,*R*) *anti* and (*S*,*S*) *syn*

transition states. The authors attribute this to be the dominant NCI contact of the aforementioned transition states due to its electrostatic nature.

2. A CH/π hydrogen bond interaction between a hydrogen on cyclohexene and the phenyl group on benzaldehyde for the (*S*,*S*) *syn* transition states.

3. The (*S*,*S*) *syn* transition states exhibit a stabilizing interaction between the *ortho*-hydrogen on the phenyl group of benzaldehyde with the hydroxyl oxygen of carboxylic acid.

Below we directly assess each of these reported sources of NCI transition state stabilization using F-SAPT.

2.4.1 $NCH^{\delta+} \cdots {}^{\delta-}O{=}C$ Interactions

A geometric inspection of the transition states shows that (*S*,*S*) and (*S*,*R*) are likely to exhibit the strongest interaction between the NCH on the pyrrolidinyl fragment and the C=O on the aldehyde. This is due to the hydrogen in NCH being within 2.43 - 2.59 Å of the aldehyde oxygen for both transition states. (*R*,*R*) and (*R*,*S*) both have longer contact distances, from 3.10 - 3.22 Å, which likely leads to a weaker interaction in these transition states.

Figure 6 provides a visual representation of the electrostatic interactions between the reacting molecules in two of the transition states considered, namely the (*S*,*R*) 1 and (*R*,*R*) 1 conformers. Figure 28 in the Ancillary Material displays the results for all 16 transition states. As previously mentioned, these F-SAPT figures are color coded such that red fragments are attracted to the other monomer, and blue fragments are repulsed by the other monomer; white fragments have net interaction energies near zero. Surprisingly, the hypothesized electrostatic stabilization between the $NCH^{\delta+}$ group of the pyrrolidinyl fragment and the ${}^{\delta-}O{=}C$ carbonyl group of the aldehyde is not evident in the F-SAPT visualizations. Instead, the pyrrolidinyl fragment is nearly white for the two conformers of Figure 6, and remains nearly white or slightly blue for all the other conformers in Figure 28; this indicates that the pyrrolidinyl fragment has no net electrostatic attraction to the benzaldehyde. The

Electrostatics:

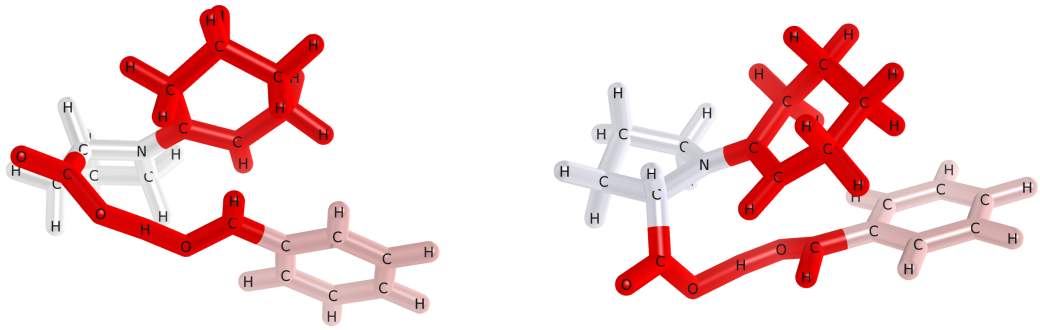


Figure 6: The electrostatic interaction between the benzaldehyde monomer and the entire enamine intermediate is visualized for (*S*,*R*) *anti* 1 (left) and (*R*,*R*) *ent-syn* 1 (right) using the terms from an F-SAPT0/jun-cc-pVDZ decomposition. Deeper red indicates a stronger attraction, and deeper blue indicates stronger repulsion.

aldehyde group is deep red for all 16 transition states, indicating a strong attraction to the enamine intermediate; however, from the color coding of the enamine intermediate, evidently the stabilizing interactions with the aldehyde group involve primarily the transferring proton from the carboxylic acid and the bond formation with the cyclohexene ring.

Figure 7 presents a breakdown of the interactions between each fragment in benzaldehyde (the aldehyde and phenyl groups) with each fragment of the enamine intermediate (the pyrrolidinyl group, the carboxylic acid group, and the cyclohexene ring). The Aldehyde $\cdots$ Pyrrolidinyl panel of Figure 7 displays the interaction energy components between the aldehyde group and the pyrrolidinyl group, and it confirms that the electrostatic interaction is repulsive (in the range of 0-2 kcal mol^{-1}) for all transition states considered.

Houk and List[3] expected the (*S*,*R*) and (*S*,*S*) transition states to be more stabilized by NCH$^{\delta+}$ $\cdots$ $^{\delta-}$O=C interactions than (*R*,*R*) and (*R*,*S*) transition states due to their closer H $\cdots$ O contacts. According to Figure 7, the (*S*,*R*) and (*S*,*S*) geometries exhibit somewhat less repulsive electrostatics in general, but there are several exceptions for particular conformers.

Given the disagreement between the F-SAPT results and the expectations from the literature regarding the favorability of the NCH$^{\delta+}$ $\cdots$ $^{\delta-}$O=C contacts, we examined these

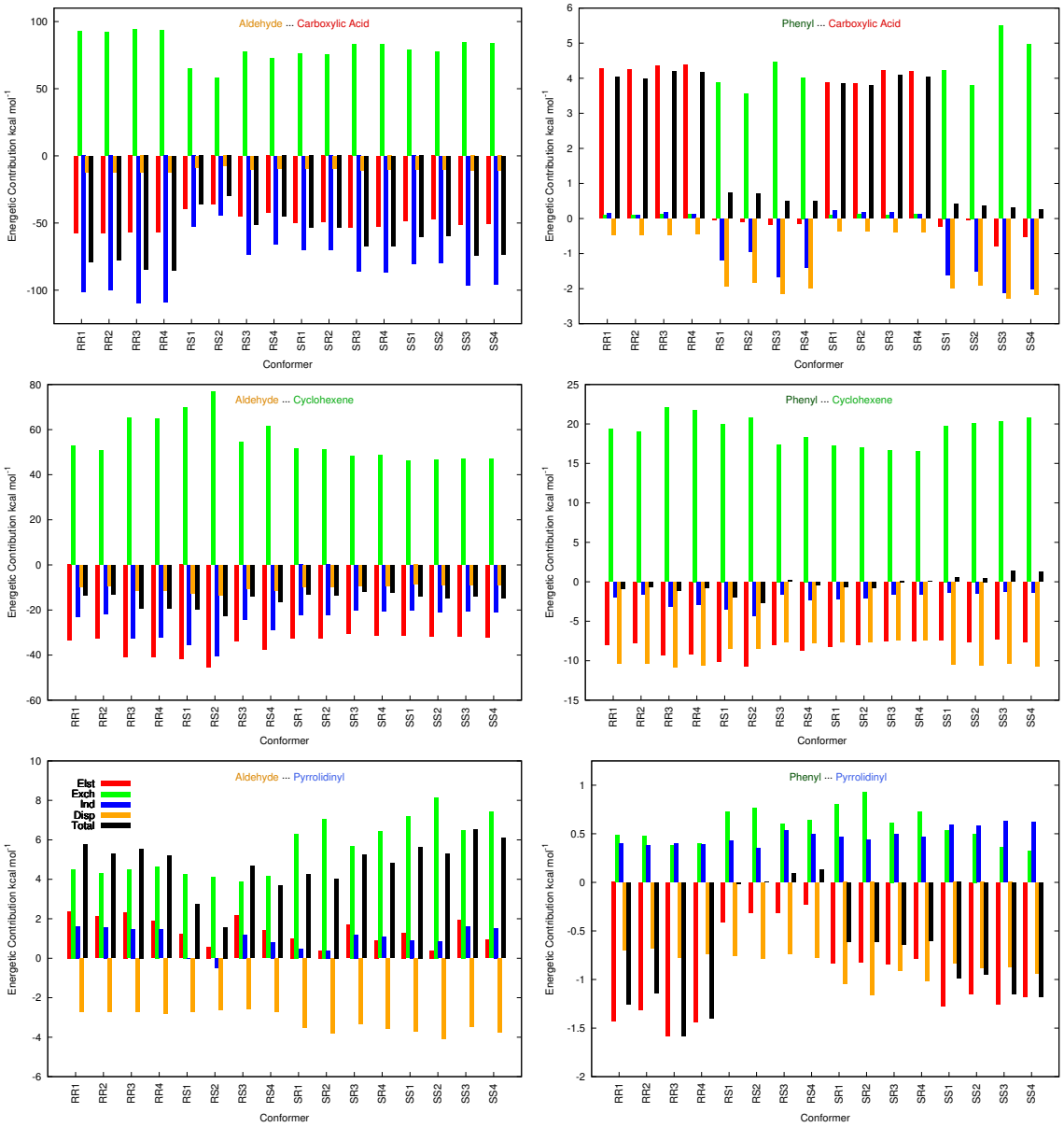


Figure 7: The F-SAPT predicted energetic components of the interaction between each of the fragments are plotted at the F-SAPT0/jun-cc-pVDZ level of theory. Only the first two letters of the stereochemical designation along with the corresponding conformer number are used to represent each possible transition state.

interactions in more detail by computing atomic charges for the transition state geometries using a natural population analysis (NPA) at the B3LYP/def2-TZVP level of theory. Atom-centered point charges are a rather crude representation of the true electrostatic interactions, but such an analysis is easier to understand. Figure 8 presents the atomic charges for the

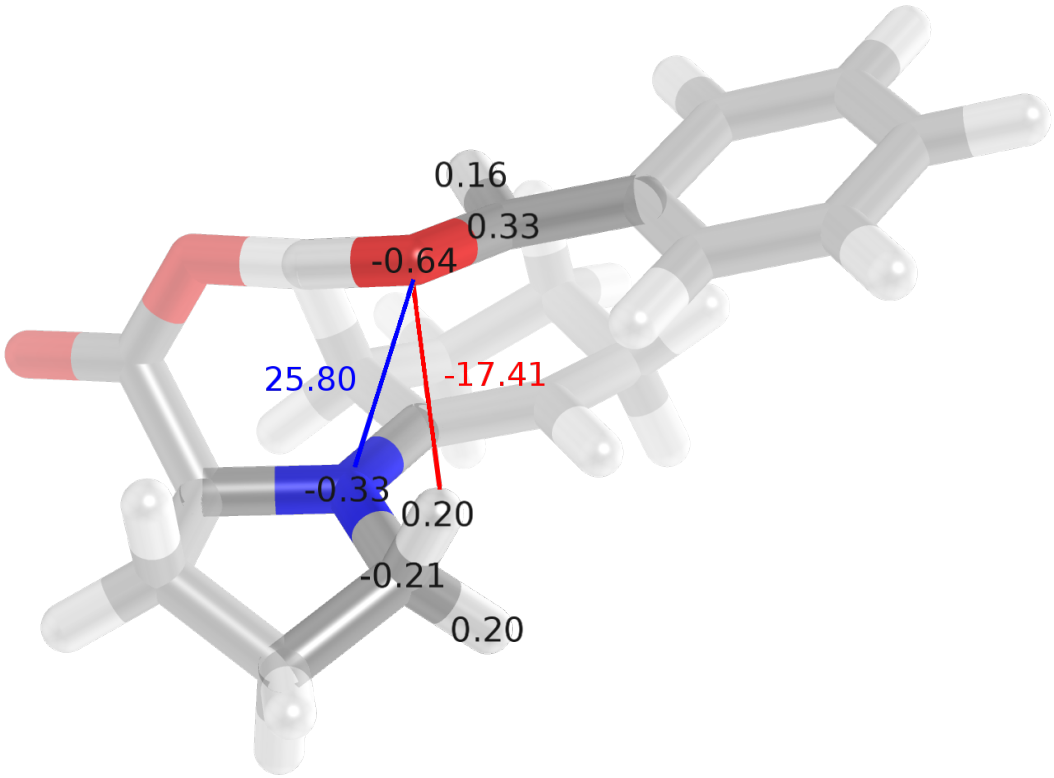


Figure 8: The primary charges involved in determining the favorability of the interaction between $NCH^{\delta+}$ of the enamine intermediate and $^{\delta-}O{=}C$ of benzaldehyde for the (*S*,*R*) *anti* 1 transition state.

most relevant atoms to the $NCH^{\delta+} \cdots {}^{\delta-}O{=}C$ contact for the (*S*,*R*) 1 transition state; results for the other transition states are similar, and a summary of the Coulomb interactions for all transition states is presented in Figures 33 – 36 of the Ancillary Material.

Using the NPA charges, the Coulomb interaction between the nearest hydrogen of the NCH_2 group and the oxygen of the aldehyde group is, as expected, strongly attractive ($\sim$ 17-19 kcal mol^{-1} for the geometries with a close contact, and $\sim$ 14-16 kcal mol^{-1} for geometries

with a farther contact). However, the interaction between the pyrrolidinyl nitrogen and the aldehyde oxygen is repulsive and of even greater magnitude ($\sim$ 24-26 kcal mol^{-1}). In the point-charge model, next most attractive contacts are (1) the more distant hydrogen of the NCH_2 interacting with the aldehyde oxygen ($\sim$ 9-12 kcal mol^{-1}), and (2) the negatively charged carbon of NCH_2 interacting with the positively charged aldehyde carbon ($\sim$ 5-7 kcal mol^{-1}). These favorable contacts are canceled out by the interaction of the aldehyde oxygen with the negative NCH_2 carbon ($\sim$ 13-15 kcal mol^{-1}) and by the repulsion between the positive aldehyde carbon and the positive NCH_2 hydrogens ($\sim$ 4 to 8 kcal mol^{-1} per contact). Other attractive interactions (N $\cdots$ H, C $\cdots$ H, C $\cdots$ C) are much weaker and are not sufficient to overcome the strength of the repulsive interactions.

Thus the F-SAPT analysis does not confirm the hypothesized electrostatic stabilization due to $NCH^{\delta+} \cdots {}^{\delta-}O{=}C$ contacts. Instead, the electrostatic interaction between the aldehyde group and the pyrrolidinyl group is found to be slightly repulsive. (This remains true if we perform alternative F-SAPT computations in which we isolate just the NCH_2 group of the pyrrolidinyl ring, instead of treating the pyrrolidinyl group as a whole). A simple analysis using atom-centered charges shows that although the $H^{\delta+} \cdots O^{\delta-}$ interaction is attractive as expected, the nitrogen is sufficiently negative that the $N^{\delta-} \cdots O^{\delta-}$ overcomes this attraction, and the overall electrostatics are repulsive.

Having analyzed the electrostatics of the aldehyde $\cdots$ pyrrolidinyl interaction in detail, let us consider the other fundamental components of the interaction. In Figure 7 we see that the most attractive interaction between these groups is not electrostatics, but London dispersion forces ($\sim$ 2-4 kcal mol^{-1}). Dispersion is somewhat more favorable for the (S,R) and (S,S) transition states than for the other stereoisomers. This is presumably a result of the closer NCH $\cdots$ O=C contacts in these geometries. On the other hand, these geometries also feature larger exchange-repulsion, so that the overall aldehyde $\cdots$ pyrrolidinyl interaction is not more favorable for these transition states. Induction is weakly destabilizing or near zero for most of the transition states. (The overall induction term must be attractive instead of repulsive, but when we look at any pair of fragments in F-SAPT, the contribution from that particular pair can be repulsive; this just means that the polarization of

each monomer in response to the other monomer leads to a new electron distribution that is overall more favorable, but from the point of view of certain functional groups, the local interaction might become less favorable).

The total aldehyde $\cdots$ pyrrolidinyl interaction energies are overall repulsive, and roughly similar across all transition states except for (*R*,*S*) 1 and (*R*,*S*) 2, which are noticeably less destabilized than the others. These two conformers feature among the least destabilizing exchange-repulsion and electrostatic terms, perhaps because they are among those with the *longer* NCH $\cdots$ O=C contacts. The relative favorability of these two conformers for the aldehyde $\cdots$ pyrrolidinyl part of the NCI between the reactants does not correlate with the experimentally observed preference for the (*S*,*R*) and (*S*,*S*) products.

2.4.2 C–H/π Interaction

Dispersion:

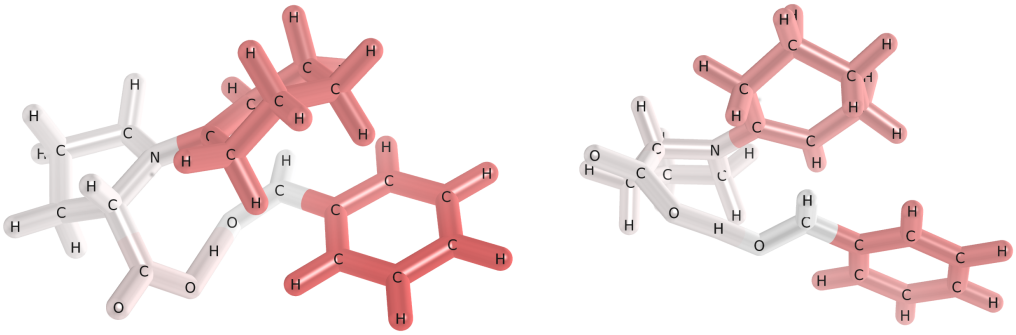


Figure 9: The dispersion interaction between the phenyl group and the entire enamine intermediate is visualized for (*S*,*S*) *syn* 1 (left) and (*S*,*R*) *anti* 1 (right) using our F-SAPT0/jun-cc-pVDZ decomposition. In this figure, aldehyde interactions are excluded.

Attractive C–H/π interactions between an aliphatic C–H group and an aromatic π system have been observed in many systems,[47] and Krenske and Houk have argued that they may control the stereochemical outcome of numerous addition reactions involving aromatic substituents.[38] These interactions are typically stabilizing by $\sim$ 1 to 5 kcal mol^{-1},

and previous work using SAPT suggests[59] that this stabilization arises from two sources: (1) London dispersion forces, and (2) an electrostatic attraction between the partial positive charge on the hydrogen and the negatively charged π face of the aromatic. Dispersion contributions are about 1.6x larger than electrostatic contributions for the prototype CH_4 $\cdots$ benzene interaction.[59]

Rzepa and co-workers[2] claim that C–H/π interactions lead to large dispersion stabilization for the (*S*,*S*) transition states. Indeed, geometric analysis of these transition states shows a favorable C–H/π contact geometry; the nearest C–H bond is directly over and oriented towards the π cloud of the phenyl group, and with contact distances of 2.58 - 2.65 Å between the nearest hydrogen of cyclohexene and the center of mass of the carbons of the phenyl ring. This favorable contact could help explain the experimental preference for (*S*,*S*) products, although not the nearly equal amount of (*S*,*R*) products. However, geometric analysis also shows similarly favorable C–H/π interactions (2.58 - 2.72 Å separation) for the (*R*,*R*) transition states, which do not lead to an experimentally favored product. For the other transition state stereoisomers, the C–H/π arrangement is not as favorable, and the contact distances are typically larger, 3.21 - 3.38 Å.

The presence of favorable C–H/π contacts between the phenyl and cyclohexene groups is also supported by the dispersion and electrostatic components of our F-SAPT analysis. Figure 9 illustrates a stronger dispersion interaction between cyclohexene and the phenyl group for the (*S*,*S*) 1 geometry, which features a close C–H/π contact, than in the (*S*,*R*) 1 transition state, which does not. (For simplicity, this figure shows the F-SAPT dispersion only between the phenyl group and the enamine intermediate, with the aldehyde contributions suppressed.)

The Phenyl $\cdots$ Cyclohexene panel of Figure 7 displays each component of the phenyl $\cdots$ cyclohexene interaction for all 16 transition states, and it confirms that the geometries with the closer C–H/π contacts, (*S*,*S*) and (*R*,*R*), have greater dispersive stabilization ($\sim$ 10-11 kcal mol^{-1}) than the (*R*,*S*) and (*S*,*R*) geometries ($\sim$ 7-8 kcal mol^{-1}). All transition states have similar, stabilizing electrostatic contributions ($\sim$ 7-8 kcal mol^{-1}) except for (*R*,*R*) 3, (*R*,*R*) 4, (*R*,*S*) 1, and (*R*,*S*) 2 ($\sim$ 9-10 kcal mol^{-1}). In all cases where a favorable

C–H/π interaction is expected, the dispersion contribution is 1.2 to 1.4 times greater than electrostatics, roughly in accordance with the 1.6 ratio for $CH_4 \cdots$ benzene.[59] Induction contributions are stabilizing or close to zero for all transition states.

Thus, for the attractive components of the interaction, there seems to be a preference for the (*S*,*S*) and (*R*,*R*) transition state stereoisomers, consistent with their more favorable C–H/π geometries, although the (*R*,*S*) 1 and (*R*,*S*) 2 conformers are also rather favorable due to their having the most attractive electrostatic interactions. Unfortunately, however, the transition states with the most favorable attractive components are not necessarily the ones with the most favorable *total* interaction energies. For the phenyl $\cdots$ cyclohexene interactions, Figure 7 shows that the transition states with the most favorable attractions also tend to have the most unfavorable exchange-repulsion terms. When all the contributions are added to yield the total non-covalent interaction energies, the (*R*,*S*) 1 and (*R*,*S*) 2 transition states are most stabilized ($\sim$ 2-3 kcal mol^{-1}) while (*S*,*S*) 3 and (*S*,*S*) 4 are the most destabilized ($\sim$ 1 kcal mol^{-1}). All other transition states are either stabilized or destabilized by about 1 kcal mol^{-1} or less. Unfortunately, and perhaps surprisingly, these results for the total interaction energies between the phenyl and cyclohexene groups do not correlate with the presence or absence of good C–H/π contacts.

2.4.3 Phenyl *ortho*-Hydrogen $\cdots$ Carboxylic Acid

A geometric inspection of the transition states for the *ortho*-hydrogen $\cdots$ carboxylic acid contact reveals contact distances of 2.18-2.32 Å for (*S*,*S*) and (*R*,*S*) transition states. In these transition states, the *ortho*-hydrogen is oriented towards an oxygen in the carboxylic acid, while the (*S*,*R*) and (*R*,*R*) transition states feature an *ortho*-hydrogen that is neither close ($\sim$ 4.9 Å) nor oriented correctly. Therefore, we expect to see non-covalent stabilization through electrostatics and dispersion for (*S*,*S*) and (*R*,*S*) transition states but not (*S*,*R*) and (*R*,*R*).

An F-SAPT illustration of this interaction is presented in Figure 10; aldehyde interactions have been suppressed for simplicity. The (*S*,*S*) transition state pictured has a red

Electrostatics:

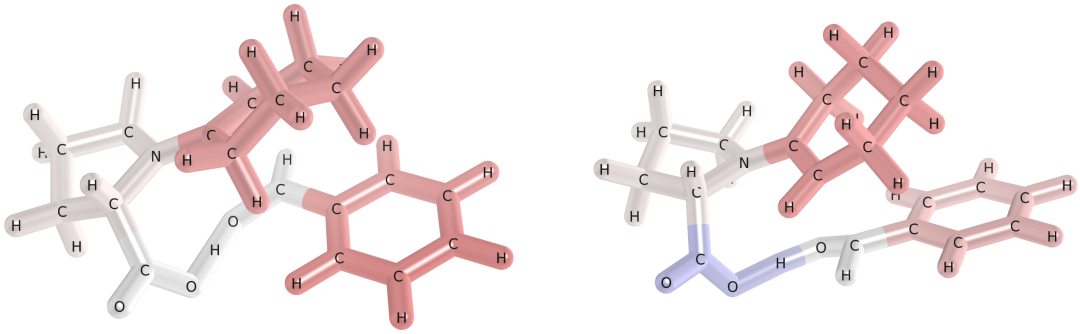


Figure 10: The electrostatic interaction between the phenyl group and the enamine intermediate is visualized for (*S*,*S*) *syn* 1 (left) and (*R*,*R*) *ent-syn* 1 (right) at the F-SAPT0/jun-cc-pVDZ level of theory. In this figure, aldehyde interactions are excluded.

phenyl group, indicating that it interacts favorably with the enamine intermediate. A significant part of this favorable electrostatic interaction is actually due to the cyclohexyl group, as indicated by its red color. The carboxylic acid group is nearly white, indicating little net electrostatic attraction between it and the phenyl group. However, in the (*R*,*R*) transition state pictured in the right of Figure 10, the phenyl group is less red (less favorable interactions with the enamine intermediate) and the carboxylic acid group is now slightly blue, indicating an unfavorable electrostatic interaction with the phenyl group. Hence, the close contact and favorable alignment between the phenyl *ortho*-hydrogen and the carboxylic oxygen allows for a favorable electrostatic contact that compensates for an otherwise unfavorable phenyl $\cdots$ carboxylic interaction.

More detail is provided by Figure 7, which presents all the components for the phenyl $\cdots$ carboxylic interaction for all 16 transition state conformers. The Figure shows a dramatic difference between the (*R*,*S*) and (*S*,*S*) transition states vs. the (*S*,*R*) and (*R*,*R*) transition states. The (*R*,*S*) and (*S*,*S*) geometries experience significant dispersion stabilization ($\sim$ 2 kcal mol^{-1}), whereas this stabilization is about 0.5 kcal mol^{-1} or less for other geometries. Likewise, (*R*,*S*) and (*S*,*S*) geometries feature significant stabilizing induction contributions

(1-2 kcal mol^{-1}), whereas induction is negligible for other geometries. The electrostatic contributions for (*S*,*R*) and (*R*,*R*) geometries are destabilizing by about 4 kcal mol^{-1}, whereas they are about zero or weakly stabilizing for (*R*,*S*) and (*S*,*S*) geometries. The significantly more favorable attractive interactions between the phenyl and carboxylic acid groups in the (*R*,*S*) and (*S*,*S*) transition states are opposed by larger exchange-repulsion contributions, which are typically around 4-5 kcal mol^{-1}; exchange-repulsion is nearly negligible for the (*S*,*R*) and (*R*,*R*) geometries. Nevertheless, the overall interaction energies between the phenyl and carboxylic acid groups are substantially more favorable for the (*R*,*S*) and (*S*,*S*) transition states, which feature close contacts between the phenyl *ortho*-hydrogen and the hydroxyl oxygen of the carboxylic acid.

2.4.4 Summary of Non-Covalent Interactions and Comparison to Experiment

A seemingly favorable NCH$^{\delta+}$ $\cdots$ $^{\delta-}$O=C interaction, identified by geometric analysis[3] and apparently confirmed by electron-density-based analysis[2] using the NCIPLOT program,[13] exists in the (*S*,*R*) and (*S*,*S*) transition states, which correspond to the experimentally favored stereoisomer products. However, in this case apparently the correlation is fortuitous, because direct computation of the interaction between the aldehyde and pyrrolidinyl fragments yields repulsive interaction energies, with similar values for all transition states except for smaller destabilization for the (*R*,*S*) 1 and (*R*,*S*) 2 conformers. The expected electrostatic attraction between the NCH$^{\delta+}$ and $^{\delta-}$O=C moieties is found to actually be repulsive due to the large partial negative charge on the nitrogen.

An aromatic CH/π interaction between a hydrogen on the cyclohexene ring and the phenyl group of benzaldehyde was noted[2] for the (*S*,*S*) transition states, which correspond to one of the two favored products. However, this contact also exists for the unfavored (*R*,*R*) transition states. The dispersion component of the F-SAPT analysis shows a preference for these stereoisomers, while the electrostatic interaction is similar for all stereoisomers except for a preference for two (*R*,*R*) and two (*R*,*S*) conformers. However, when all SAPT components are added up, the overall interaction energies between the cyclohexene and phenyl groups are actually more favorable for the (*R*,*S*) 1 and (*R*,*S*) 2 conformers, which

lack the proposed CH/π interaction.

Finally, density-based analysis with NCIPLOT also suggested favorable stabilizing interactions between an *ortho*-hydrogen of the phenyl ring and the hydroxyl oxygen of the carboxylic acid.[2] This interaction is present in the (*S*,*S*) and (*R*,*S*) transition states. F-SAPT analysis clearly supports the favorability of these contacts, which lead to substantially more favorable non-covalent interactions between the phenyl and carboxylic acid groups in these stereoisomers.

Thus, out of three hypothesized stabilizing interactions, one turns out to be destabilizing, another turns out to be canceled by exchange-repulsion forces, and only the third is confirmed by direct computation. The overall interaction between the pyrrolidinyl and aldehyde groups favors the (*R*,*S*) 1 and (*R*,*S*) 2 conformers, as does the overall interaction between the cyclohexene and phenyl groups. Unfortunately, these are not the experimentally favored stereoisomers of the products. The overall interaction between the phenyl and carboxylic acid groups favors the (*S*,*S*) and (*R*,*S*) stereoisomers. While the (*S*,*S*) stereoisomers are one of the favored products experimentally, the (*R*,*S*) isomers are not. Thus, none of the hypothesized non-covalent interactions seems to control the stereoselectivity.

Figure 11 presents the total SAPT interaction energies between the two reacting molecules, summed over all interacting pairs of fragments. The total SAPT interaction energies are similar for many of the transition states, but with a noticeable preference for the (*R*,*R*) stereoisomers and a less pronounced preference for the (*S*,*S*) 3 and (*S*,*S*) 4 conformers. The stronger stabilizing interaction energies for the (*R*,*R*) stereoisomers are primarily a result of a preference for these stereoisomers in the very strong interaction between the reacting aldehyde and carboxylic acid groups (see Figure 7). We note that the *total* SAPT interaction energies also fail to correlate well with the experimentally observed preference for (*S*,*R*) and (*S*,*S*) products. This strongly suggests that stereoselectivity in this reaction is not solely governed by non-covalent interactions, but is also substantially influenced by deformation energies (the other half of the "distortion-interaction" model of transition states[20] mentioned in the introduction).

Our primary goal in this paper has been to directly analyze the proposed non-covalent

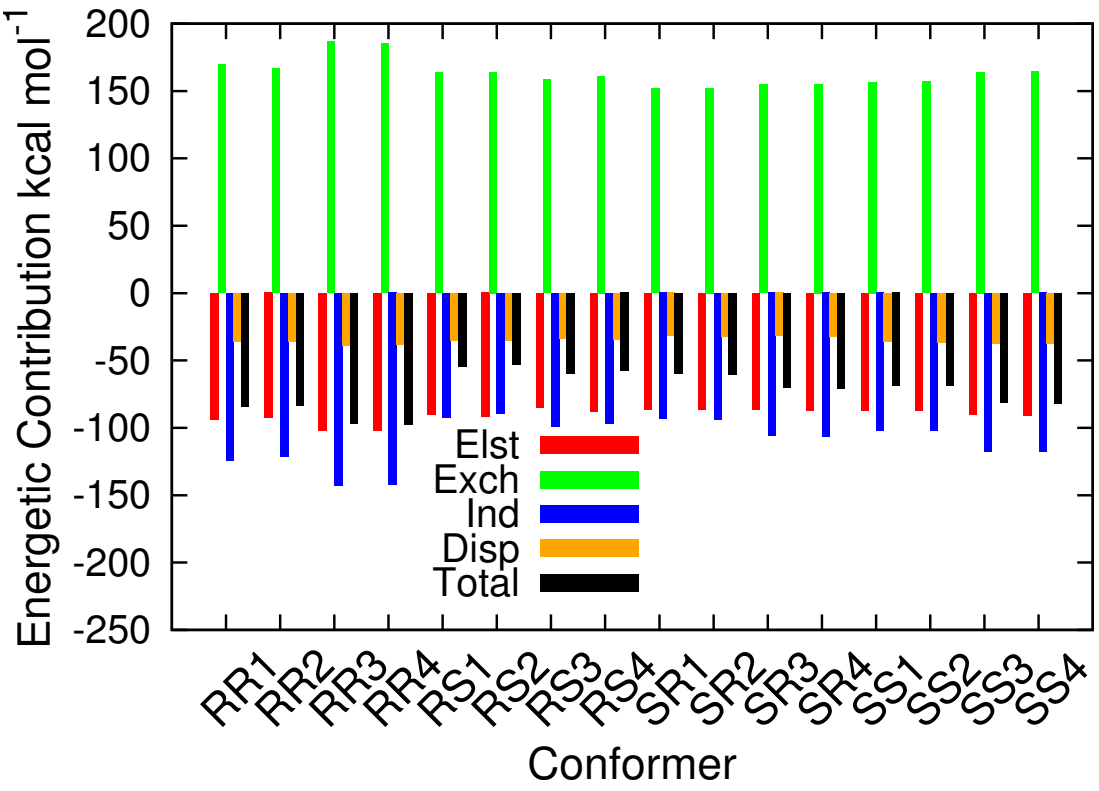


Figure 11: The total SAPT decomposition of the energetics in each transition state is plotted at the SAPT0/jun-cc-pVDZ level of theory. This plot only accounts for total monomer interactions, meaning the energetics are for the whole enamine intermediate interacting with benzaldehyde.

interactions in the Houk-List model of proline-catalyzed aldol reactions. However, given that we have computed CCSD(T)/CBS energies for the transition states, it is interesting to explore whether these higher-level results exhibit better agreement with experimental product ratios than the B3LYP-D3/TZVP/SCRF=DMSO results of Rzepa and co-workers.[2] That study predicted a product ratio of 99.3 (*S*,*R*) : 0.7 (*S*,*S*), compared to an experimental[3] ratio of 45-47 : 43-45. Table 1 presents the relative electronic energies of the transition states computed at the B3LYP-D3/TZVP/SCRF=DMSO level of theory (ΔE_{solv}) and the corresponding relative Gibbs free energies at 298K ($\Delta\Delta G_{\text{solv}}$). We then present results using our gas-phase CCSD(T)/CBS estimates for the transition state energies (ΔE_{gas}). We were unable to perform solvent corrections and/or thermodynamic corrections to our bare

		CCSD(T)/CBS			B3LYP-D3/TZVP			Experiment
Isomer	Conf.	ΔE_{gas}	$\Delta\Delta G_{\text{solv}}$	% Pop.	ΔE_{solv}	$\Delta\Delta G_{\text{solv}}$	% Pop.	% Pop.
(*S*,*R*) *anti*	1	0.00	0.00	98.81	0.00	0.00	99.30	45-47
	2	0.45	0.15		0.25	-0.05		
	3	2.77	3.45		1.47	2.15		
	4	2.58	3.33		1.07	1.82		
(*S*,*S*) *syn*	1	1.37	2.79	0.98	1.57	2.99	0.68	43-45
	2	1.65	2.84		1.73	2.92		
	3	3.35	5.38		2.70	4.73		
	4	2.81	5.07		1.96	4.22		
(*R*,*R*) *ent-syn*	1	7.42	9.18	0.00	5.24	7.00	0.01	5-7
	2	5.01	6.91		3.19	5.09		
	3	8.05	9.72		5.87	7.54		
	4	6.45	8.03		4.30	5.88		
(*R*,*S*) *ent-anti*	1	6.26	5.84	0.21	7.51	7.09	0.01	3-5
	2	3.37	3.35		5.35	5.33		
	3	8.13	7.60		8.36	7.83		
	4	5.57	4.90		6.43	5.76		

Table 1: The relative total energies of the transition states (ΔE), relative Gibbs free energies at 298K ($\Delta\Delta G$), and percent populations are listed for the CCSD(T)/CBS estimates obtained in this work and the previous best estimate by Rzepa and co-workers[2] at the B3LYP-D3/TZVP/SCRF=DMSO level of theory. CCSD(T)/CBS ($\Delta\Delta G$) results include solvent and thermodynamic corrections obtained at the B3LYP-D3/TZVP/SCRF=DMSO level. Experimentally measured[3] product ratios are presented for comparison.

electronic energy differences at the coupled-cluster level, so we simply applied these corrections as determined by Rzepa and co-workers at the B3LYP-D3/TZVP/SCRF=DMSO level to obtain approximate CCSD(T)/CBS $\Delta\Delta G_{\text{solv}}$ values.

The $\Delta\Delta G_{\text{solv}}$ values are fairly consistent between the B3LYP-D3/TZVP and CCSD(T)/CBS levels of theory, although differences can be as large as 2.2 kcal mol^{-1}. Nevertheless, predicted product ratios are very similar between B3LYP-D3/TZVP and CCSD(T)/CBS; both predict about 99% of (*S*,*R*) and about 1% of (*S*,*S*). Unfortunately this leaves the discrepancy with experiment unresolved. However, we note that Rzepa and co-workers did not seek any pre-reactive complexes of the two reactants; if such complexes experience significant stabilization, they might influence the relative barrier heights and hence the product ratios. Additionally, the noticeable differences in energetics between the B3LYP-D3/TZVP and CCSD(T)/CBS might be taken to indicate that the B3LYP-D3/TZVP transition state geometries might change significantly if they could be re-optimized at higher levels of theory (unfortunately, this would be rather expensive computationally).

2.5 Conclusions

The "distortion-interaction" model[20] and the "activation strain" model[21] state that transition state energetics are determined by non-covalent interaction energies and by the energy required to deform the reactants into the transition state geometry. In principle, a better understanding of non-covalent interactions could be used to reduce barrier heights and/or control stereoselectivity. The recently-developed functional-group partitioning of symmetry-adapted perturbation theory (F-SAPT)[50] provides a theoretical tool for directly computing non-covalent interactions between pairs of functional groups. The present study represents its first application to analyzing non-covalent interactions in transition states. We have studied the proline-catalyzed aldol reaction between benzaldehyde and cyclohexanone as a simple model system for which several stabilizing non-covalent interactions have been hypothesized to play a role in stereoselectivity.

Of the three previously-hypothesized stabilizing non-covalent interactions, F-SAPT confirms only one, a favorable contact between an *ortho*-hydrogen of the phenyl group and the hydroxyl oxygen of the carboxylic acid. A hypothesized CH/π interaction between a hydrogen of cyclohexene and phenyl is found to have favorable attractive terms but is canceled by unfavorable exchange-repulsion terms. A hypothesized stabilizing $NCH^{\delta+} \cdots {}^{\delta-}O{=}C$ contact thought to control stereoselectivity is found to actually be repulsive due to the negative charge on the nitrogen. These contacts were thought to be stabilizing on the basis of a geometric analysis of the transition state and chemical intuition, and/or on the basis of electron density analysis via the NCIPLOT program,[13] which relies on correlations between the properties of the electron density and what are typically found to be favorable non-covalent interactions. The reaction studied appears to provide a challenge for such geometry and/or density-based analysis.

F-SAPT is also challenged by the transition states studied; the usual accuracy[49] of the underlying SAPT0/jun-cc-pVDZ approach is degraded for the present system, no doubt because the transferring proton and the forming carbon-carbon bond lead to a very strong interaction between the monomers (the theory assumes modest to weak interactions). Nevertheless, comparison against CCSD(T)/CBS results shows that although the relative overall

SAPT0 interaction energies have errors of several kcal mol^{-1}, they demonstrate essentially the same energetic ordering as CCSD(T)/CBS, meaning that they should be reliable in predicting trends. Additionally, the non-covalent interactions hypothesized in the literature to control stereoselectivity are not between functional groups that are forming bonds, and hence F-SAPT should be more reliable for these particular pairs of functional groups.

Whether one uses F-SAPT or the high-quality CCSD(T)/CBS results, *interaction energies* between the two reactants do not correlate well with either the experimentally observed product ratios or with the differences between the CCSD(T)/CBS *total energies* of the transition states, which include geometry deformation effects. This indicates that, for the present reaction, the "distortion" terms of the "distortion-interaction" model are at least as important as the interaction terms for understanding stereoselectivity. It remains to be seen whether other classes of reactions are more completely controlled by the interaction terms.

Relative energies of the transition states computed with CCSD(T)/CBS are in reasonably good agreement with previous B3LYP-D3/TZVP results. Either approach, when corrected for solvent effects and thermodynamic corrections, estimates around a 99:1 product ratio for (S,R) : (S,S) stereoisomers, compared to experimental results showing a roughly equal proportion of these isomers. We hypothesize that better agreement with experiment may require searching for pre-reactive complexes and/or obtaining higher-quality transition state geometries.

CHAPTER III

ANALYSIS OF TRANSITION STATE STABILIZATION BY NON-COVALENT INTERACTIONS IN THE ADDITION OF ORGANOBORON REAGENTS TO FLUOROKETONES

3.1 Summary

This work seeks to apply symmetry-adapted perturbation theory (SAPT) to the recent study of Hoveyda and co-workers [K. A. Lee *et al., Nat. Chem.* 2016, **8**, 768] where an allyl addition to a ketone became enantioselective when the ketone was fluorinated. Through the application of atomic SAPT (A-SAPT) and functional-group SAPT (F-SAPT), the non-covalent interactions between specific atoms and functional groups in the transition states associated with the fluoroketone reactions can be quantified. Our A-SAPT analysis confirms that a H $\cdots$ F contact thought to enhance stereoselectivity shows a strong preference for one of the transition states leading to the experimentally observed product enantiomer. Other key atom-atom contacts invoked to rationalize relative transition state energies are also found to behave as expected based on chemical intuition and contact distances. On the other hand, hypothesized steric clashes between substrate phenyl or *ortho*-methyl phenyl groups and the catalyst are not supported by F-SAPT computations, and indeed, these are actually favorable π-π interactions.

3.2 Introduction

In the activation-strain[21] or distortion-interaction[20] models, transition state (TS) barrier heights are the sum of the deformation energy penalties the reactants must pay to adopt the TS structure and the non-covalent interaction energy between the reactants. Design of more effective catalysts would be aided by a better understanding of these contributions to the energy. Indeed, several recent studies have explored the idea of improving the selectivity of organocatalytic reactions by tuning non-covalent interactions in the transition state.[38, 43, 58, 66, 67, 82]

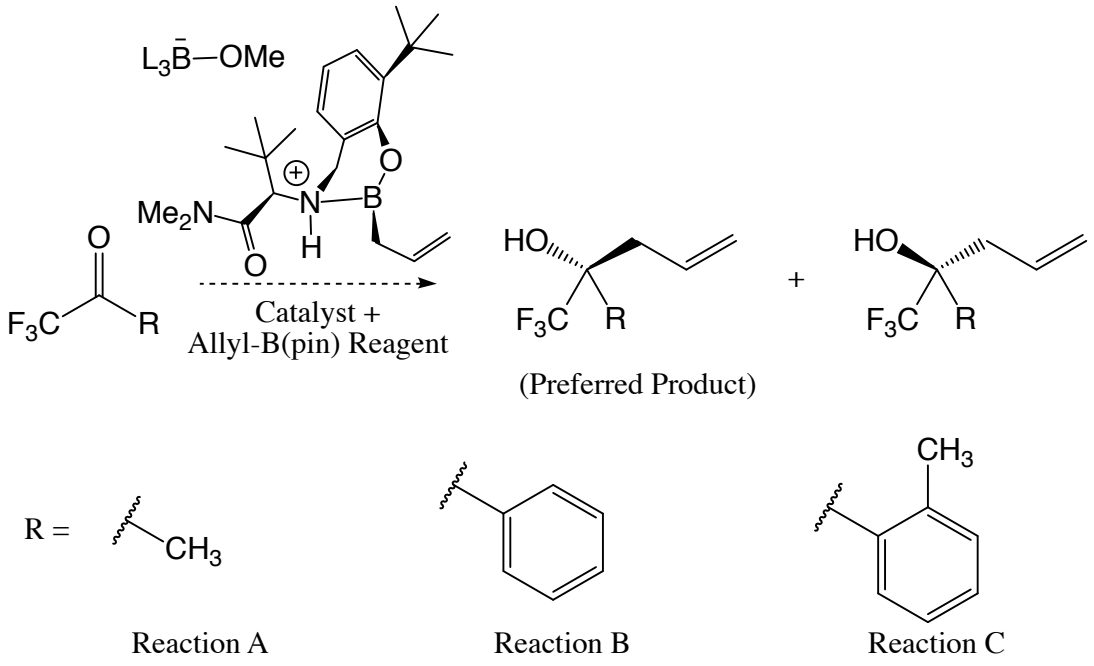


Figure 12: Enantioselective addition of an allyl group to fluoroketones, through a complex formed between the allyl-boron reagent and an aminophenol catalyst.

Recently, Hoveyda and co-workers[40] hypothesized that electrostatic interactions between a positively charged ammonium moiety and a fluorine atom on the substrate could enhance enantioselectivity in the addition of allyl and allenyl groups to fluoroketones in reactions like those in Figure 12. The catalyst is generated by reacting an aminophenol molecule with an organoboron reagent bearing the allyl or allenyl group to be added, and reactions using several different trifluoro-methyl ketone substrates were examined. Density functional theory (DFT) was used to locate possible TS structures for the simplest substrate, 1,1,1-trifluoroacetone (reaction A of Figure 12), and the lowest-energy TS was one that featured the hypothesized H $\cdots$ F interaction. When experiments were performed using a variety of trifluoroketone substrates, the desired enantiomer was obtained with a high enantiomeric excess in most cases. For example, substrate 2,2,2-trifluoroacetophenone (reaction B in Figure 12) led to an enantiomeric ratio of 96:4. This contrasts with an enantiomeric ratio of 32:68 (i.e., the other enantiomer is preferred) for the non-fluorinated substrate. Both the DFT and experimental results are consistent with the hypothesis that H $\cdots$ F interactions

help stabilize the transition states leading to the desired products.

To rationalize the relative energies of the possible TS's, Hoveyda et al.[40] invoked various additional non-covalent interactions shown in Figure 13, based on the TS geometries and chemical intuition. We recently reported[4] a quantum mechanical study on the Houk-List mechanism for intermolecular aldol additions showing that chemical intuition is not necessarily reliable for understanding how non-covalent interactions might stabilize or destabilize the transition state. The original hypothesis for the origin of the stereoselectivity in that reaction was that the TS leading to the dominant product is stabilized by a favorable $NCH^{\delta+}\cdots^{\delta-}O{=}C$ H-bonding contact. However, direct computation of the strength of this contact using functional-group symmetry-adapted perturbation theory (F-SAPT)[50] showed that this is in fact a destabilizing interaction (due to repulsion between negative partial charges on the N and O atoms), and its strength in the possible transition state structures does not correlate with the observed product ratios.

In this paper, we seek to quantify the various non-covalent interactions that Hoveyda et al.[40] suggest are important in stabilizing or destabilizing the transition state structures for the organoboron catalyzed addition of an allyl unit to a fluoroketone substrate. We use F-SAPT and also the atomic version, A-SAPT.[51] Unlike our previous study of the Houk-List mechanism, here we find that chemical intuition about the nature of these contacts is mostly supported by the A-SAPT and F-SAPT analysis.

3.3 Theoretical Methods

Symmetry-adapted perturbation theory[33] (SAPT) decomposes the interaction energy between two monomers in terms of electrostatics, exchange-repulsion, induction/polarization, and London dispersion forces.[33] Atomic SAPT (A-SAPT)[51] and functional-group SAPT (F-SAPT)[50] offer a more fine-grained analysis by partitioning densities to provide these same energetic components between interacting pairs of atoms or functional-groups of each monomer. These methods are based upon SAPT using Hartree–Fock monomer wavefunctions, with intermolecular interactions treated through second-order, and with intramolecular electron correlation neglected (sometimes abbreviated SAPT0). In this work, we use

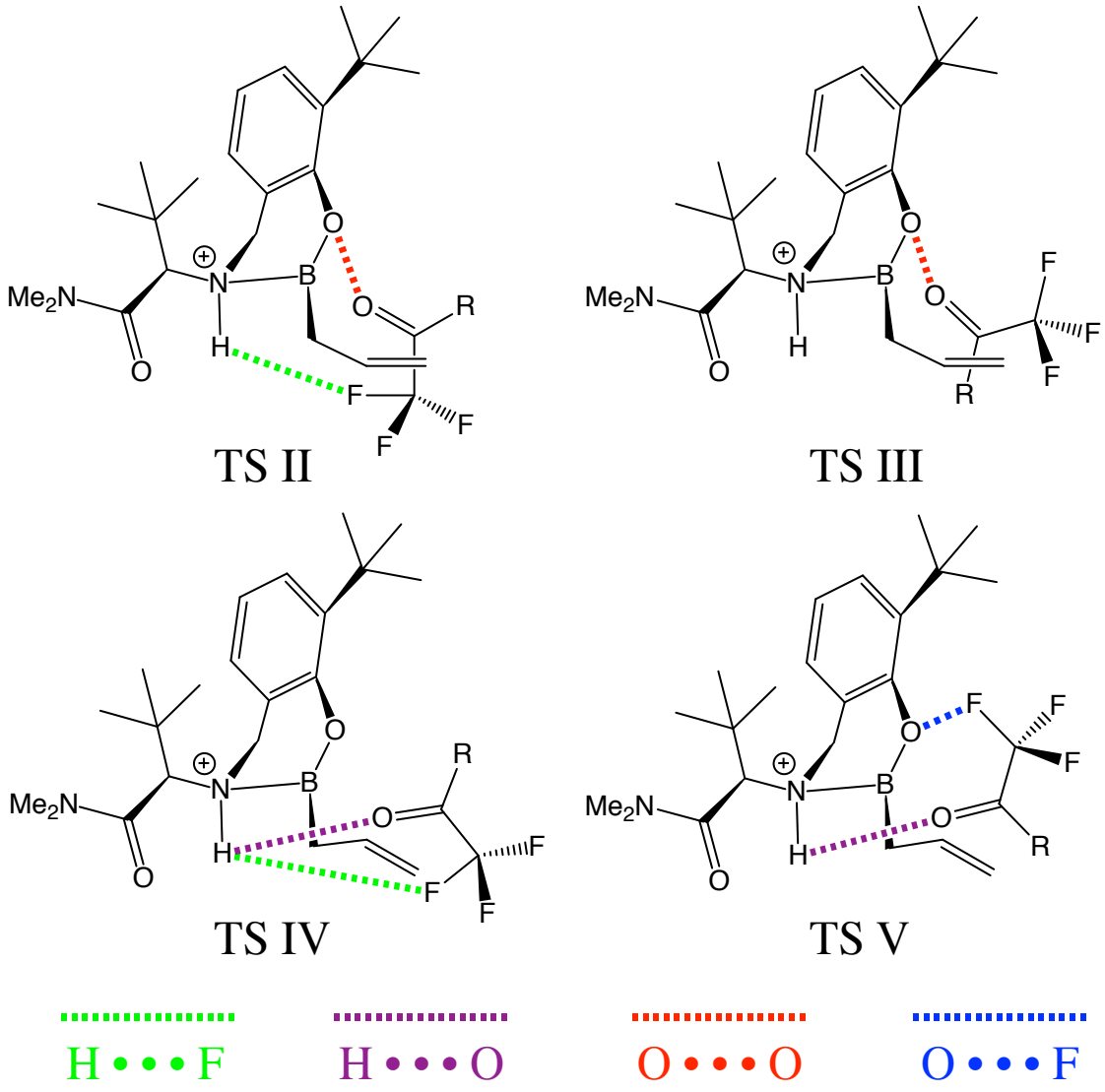


Figure 13: Transition state structures for the reaction between trifluoro-methyl ketones and organoboron catalyst complexes from Figure 12, using the numbering from Hoveyda and co-workers. Also shown are the non-covalent interactions suspected to determine the enantioselectivity of the reaction: repulsion between the oxygen of the carbonyl group of the fluoroketone and aryloxy oxygen of the catalyst complex in red; attraction between the carbonyl oxygen and the ammonium proton of the organoboron reagent in purple; attraction between a fluorine of the fluoroketone and the ammonium proton in green; repulsion of a fluorine of the fluoroketone and the aryloxy oxygen of the catalyst complex in blue.

A-SAPT and F-SAPT to probe specific interactions between monomers in the transition

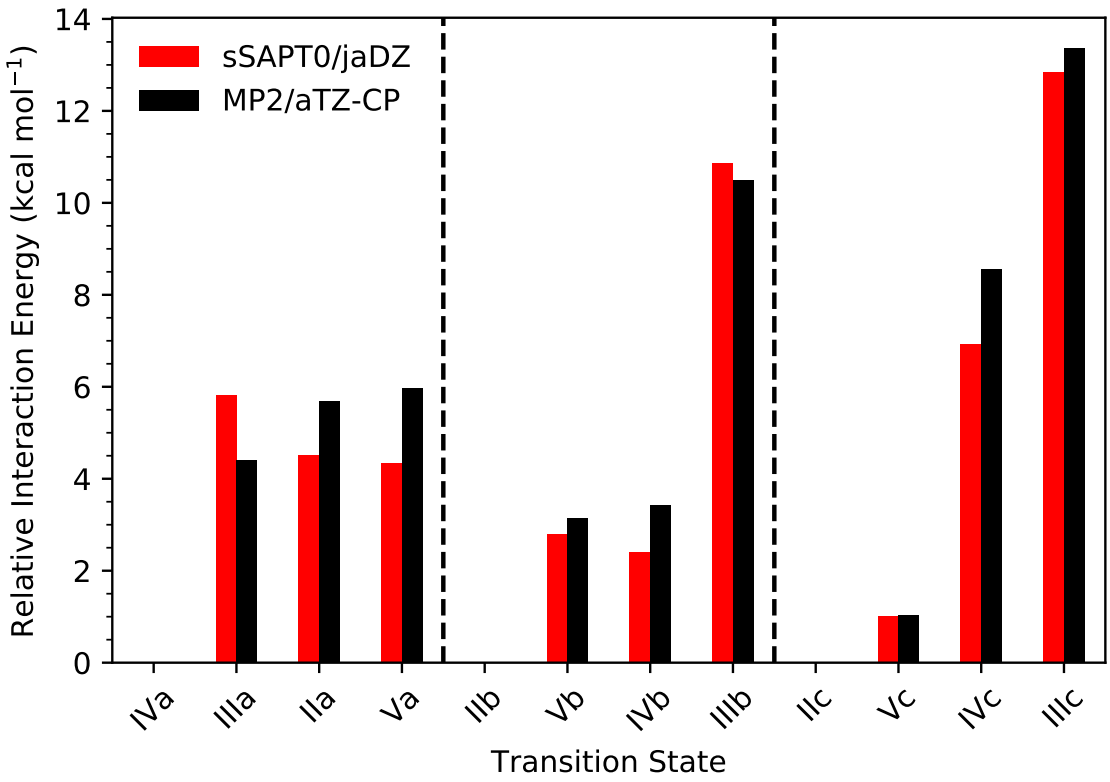


Figure 14: Interaction energies are sorted by the strength of the counterpoise-corrected MP2/aTZ estimate within each reaction family (A, B, and C) and are plotted relative to the lowest interaction energy estimated within each model chemistry (exchange-scaled sSAPT0/jaDZ and MP2/aTZ).

state. Here, the monomers will be defined as the positively charged organoboron catalyst-complex, and the neutral fluoroketone substrate. We employed DFT transition state geometries reported by Hoveyda and co-workers,[40] and we use their naming scheme, in which the transition states are labeled II-V, with II and IV leading to the experimentally favored product enantiomer (see Figure 13). A letter a-c after a TS label indicates a particular reaction, A, B, or C, as shown in Figure 12.

SAPT0 provides unusually large induction and dispersion components for closely interacting systems, such as a doubly hydrogen bonded complex or a transition state.[4,39,49] This breakdown is due to the perturbative approach underlying SAPT and also a breakdown of the single-exchange (S^2) approximation used in computing the exchange-induction and

exchange-dispersion terms.[39] In our previous study of non-covalent stabilization of transition states in proline-catalyzed aldol reactions,[4] we found that exchange-scaled SAPT (sSAPT) provided substantially improved results, consistent with its superior performance for other strongly-interacting systems like doubly hydrogen-bonded systems.[39,49] To examine the quality of the SAPT interaction energies for the systems of this study, ranging from 73 – 83 atoms, we have chosen to use counterpoise corrected (CP)[11] second-order Møller-Plesset perturbation theory (MP2),[46] with Dunning's correlation-consistent aug-cc-pVTZ basis set[17] as our benchmark. Because MP2 can overbind π-π interactions,[1] we have verified the computed interaction energies for TS's with potential for π-π interactions are within 1 kcal mol^{-1} of the CP-corrected ωB97X-V/aug-cc-pVTZ level of theory, which has been shown to produce accurate interaction energies for a variety of complexes[12] (see Figure 38 of the Ancillary Material for details).

As in our previous study applying F-SAPT to transition states,[4] exchange-scaling generally improves relative interaction energies, as seen in Figure 40 of the Ancillary Material. With exchange-scaling, sSAPT relative and absolute interaction energies differ by at most 1.9 kcal mol^{-1} and 7.9 kcal mol^{-1}, respectively, from CP-corrected MP2/aTZ results for the systems studied, which have large total interaction energies in the range 37-60 kcal mol^{-1}. For reaction A, sSAPT and MP2 agree that TS IVa is stabilized by 4-6 kcal mol^{-1} compared to the others, and that the other TS's have similar interaction energies (although MP2 and sSAPT disagree about whether IIIa is slightly stabilized or slightly destabilized compared to IIa and Va). For reaction B, sSAPT and MP2 agree that IIb is the most stable, followed by Vb and IVb at about 2 kcal mol^{-1} less stable, followed by IIIb which is much less stable. For reaction C, MP2 and sSAPT agree that IIc is the most stable, followed closely by Vc, with IVc being significantly less stable, and with IIIc being even less stable. An exchange-scaled version of A-SAPT has been implemented in a developer's version of Psi4 as described previously for F-SAPT.[4] It must be noted that the energetic quantities provided by SAPT or its A-SAPT or F-SAPT variants, e.g., the electrostatic interaction between a substrate fluorine and a hydrogen of the catalyst, cannot be directly observed by experiment; nevertheless, this type of analysis has been shown to effectively illuminate

interaction motifs, such as hydrogen bonding and C–H/π interactions.[4,50–52]

3.4 Results and Discussion

Hoveyda and co-workers obtained TS's for several substrates, including reactions A-C from Figure 12. Transition states II and IV lead to the preferred enantiomer of the product, while transition states III and V lead to the opposite enantiomer. For reaction A, IV is the lowest-lying TS, with II and V lying 2.6–3.7 kcal mol^{-1} higher in energy [depending on the density functional, either M06-2X or ωB97X-D with a 6-311++G(2df,2pd) basis]. III lies higher still, at 5.5 kcal mol^{-1} above IV. Reaction B, with R = phenyl, features similar energetics, but II and V are slightly closer in energy to IV (lying 1.5–3.2 kcal mol^{-1} higher), and III is slightly destabilized (7.6–8.3 kcal mol^{-1} above IV). For reaction C, II and IV are closer to each other (within 1.2 kcal mol^{-1}), and which is lower depends on the functional used. V lies slightly higher at 1.4–1.8 kcal mol^{-1}, while III lies substantially higher (7.1–8.3 kcal mol^{-1}). In all cases, either transition state IV or II, both of which lead to the preferred product, are found to be the lowest in energy.

These TS relative energies are rationalized by Hoveyda and co-workers primarily on the basis of various stabilizing or destabilizing non-covalent interactions (although of course distortion energies also play a role in determining the total TS energies). The key H $\cdots$ F interaction thought to engender enantioselectivity (in green, Figure 13) is hypothesized to stabilize both II and IV, which both lead to the preferred product. This contact should be stronger in II than IV because the distance between the H and F atoms is $\sim$ 2.0 Å in II, but $\sim$ 4.1 Å in IV. However, IV lies energetically below II for reactions A and B, and possibly for reaction C (depending on the functional used). Hoveyda and co-workers hypothesize that II is destabilized relative to IV because of an electrostatic repulsion between the carbonyl oxygen and the aryloxy oxygen of the catalyst complex (O $\cdots$ O interaction in red, Figure 13).

Hoveyda and co-workers hypothesize that TS's III and V lie higher in energy than II and IV because they lack the stabilizing H $\cdots$ F contact, and also because they feature repulsive interactions between the aryloxy oxygen of the catalyst complex and either the

carbonyl oxygen (O $\cdots$ O contact in red in Figure 13, for III) or a fluorine (O $\cdots$ F contact in blue in Figure 13, for V). Finally, the transition states are thought to be stabilized by a favorable electrostatic contact between the carbonyl oxygen and the ammonium proton (H $\cdots$ O contact in purple in Figure 13). This contact is expected to be more favorable in TS's V and IV than in II and III, due to a shorter H $\cdots$ O distance, and this is expected to be one reason why V lies lower in energy than III.

In addition to atom-atom contacts, steric interactions between the organoboron catalyst and bulkier R-groups (benzene and *o*-methylbenzene) of the substrate are expected to play a role in altering the enantioselectivity of reactions B and C. As a result of the steric clashing, the substrate is expected to deform into a higher energy geometry in the transition state as well. Hoveyda and co-workers argue that these two effects combine to reduce the preference for TS IV in reaction B and completely eliminate the preference for TS IV in reaction C (with TS II becoming essentially isoenergetic).[40]

In the following discussion, we will analyze the aforementioned interactions hypothesized by Hoveyda and co-workers to determine the enantioselectivity of reactions A-C. For the atom-atom interactions, this will be done within the A-SAPT model, and functional group interactions will be assessed using the F-SAPT model. Figures 15-17 show the A-SAPT predicted energetics for the interactions under study in Reactions A-C, in terms of the usual SAPT decomposition: electrostatics, exchange-repulsion, induction/polarization, and London dispersion.

3.4.1 H $\cdots$ F

Hoveyda and co-workers note that the H $\cdots$ F interaction is most pronounced in TS II by virtue of it having the closest contact ($\sim$ 2.0 Å), but is also close enough ($\sim$ 4.1 Å) to exhibit a substantial stabilizing effect in IV. Both of these TS's lead to the preferred product enantiomer.

As seen in Figures 15 – 17, A-SAPT indicates that the H $\cdots$ F total interaction is indeed most stabilizing in II, followed by IV, as predicted by Hoveyda and co-workers. This interaction is dominated by electrostatics in all TS's. In fact, II is the only TS to

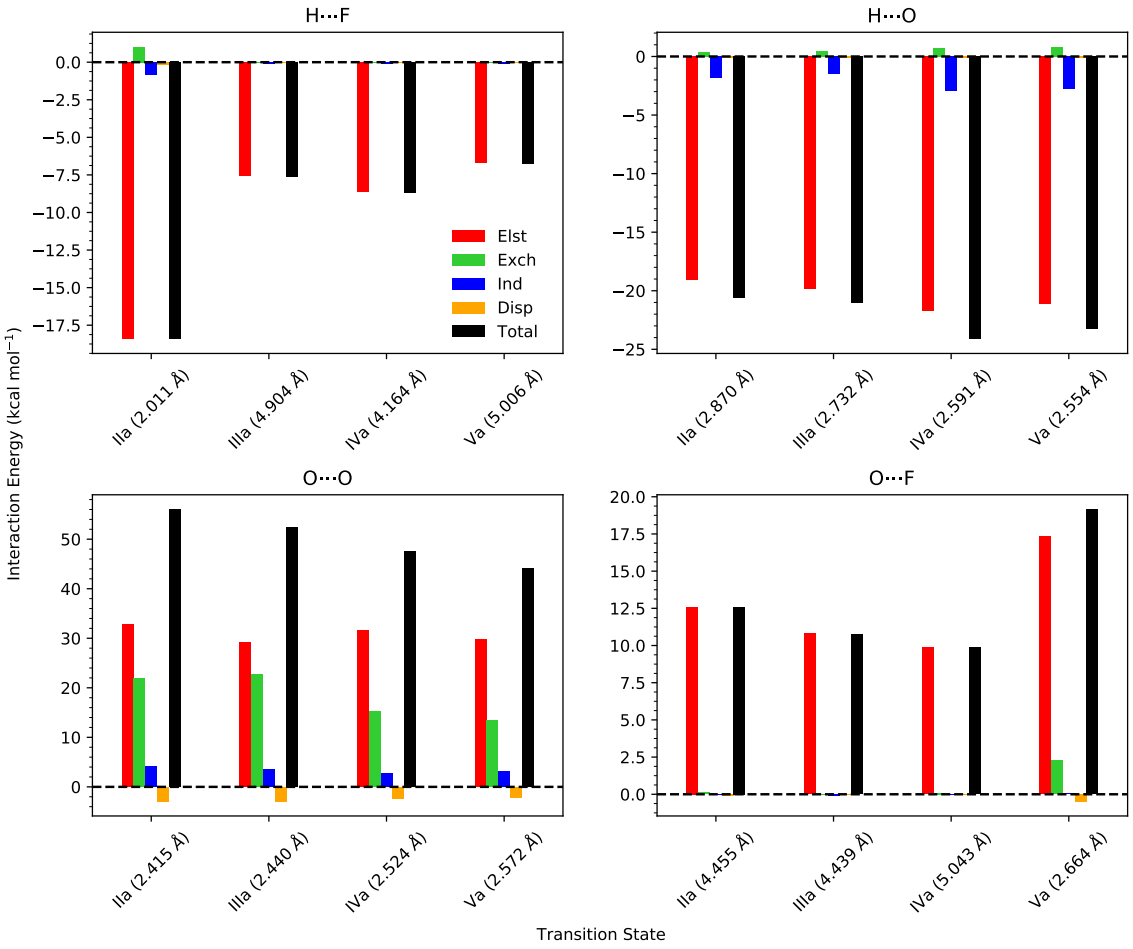


Figure 15: Exchange-scaled A-SAPT/jaDZ analysis of the interactions in transition states for Reaction A in terms of electrostatics, exchange-repulsion, induction, and dispersion. Each panel corresponds to an atom-atom interaction depicted in Figure 13.

have noticeable contributions from components other than electrostatics. With respect to electrostatics, II is stabilized by 18 – 20 kcal mol^{-1} and IV is stabilized by 7 – 9 kcal mol^{-1} across reactions A-C, making II the most stabilized from this interaction by far, as expected from having the closest contact distance by more than a factor of two. III and V are slightly less stabilized by this contact than IV, due to their larger contact distances (0.7 – 0.8 Å longer). In this case, A-SAPT confirms chemical intuition that the energetics of this contact should correlate with distance. Although II has the shortest distance and thus greatest stabilization due to this contact, Figure 14 indicates that IV is overall more stabilized by non-covalent interactions in the case of reaction A. This indicates that other

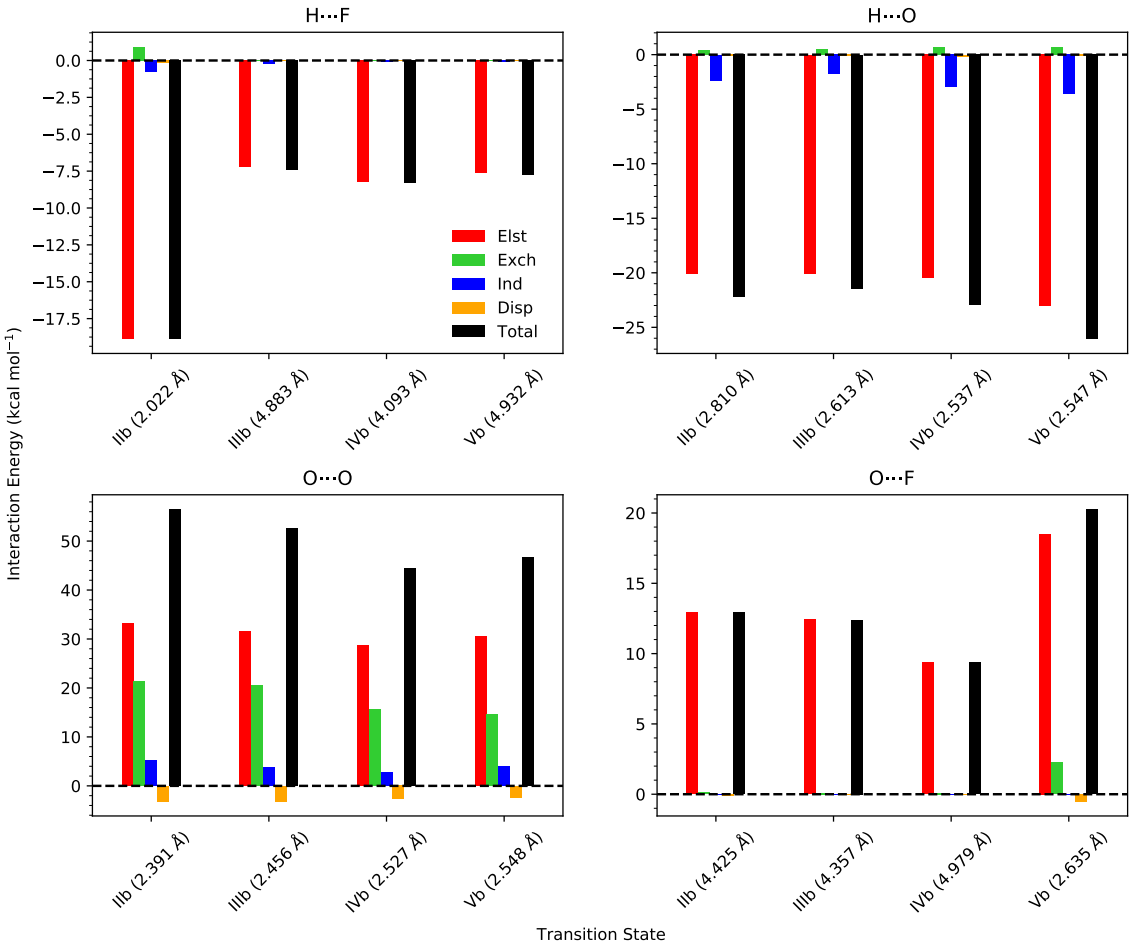


Figure 16: Exchange-scaled A-SAPT/jaDZ analysis of the interactions in transition states for Reaction B in terms of electrostatics, exchange-repulsion, induction, and dispersion. Each panel corresponds to an atom-atom interaction depicted in Figure 13.

contacts are also significant in determining the TS energetics.

3.4.2 H ··· O

Hoveyda and co-workers imply that TS's IV and V should exhibit the strongest H ··· O interaction because the chair conformation allows for a shorter contact distance. The authors suggest that this contact is one of the reasons that IV lies energetically below II. In all TS's of reactions A-C, the electrostatics component of the interaction is dominant and stabilizing by 19 – 23 kcal mol^{-1}, while induction is a slightly stabilizing component (1 – 3 kcal mol^{-1}). The exchange-repulsion component is slightly destabilizing ($<$ 1 kcal mol^{-1}), and dispersion is negligible.

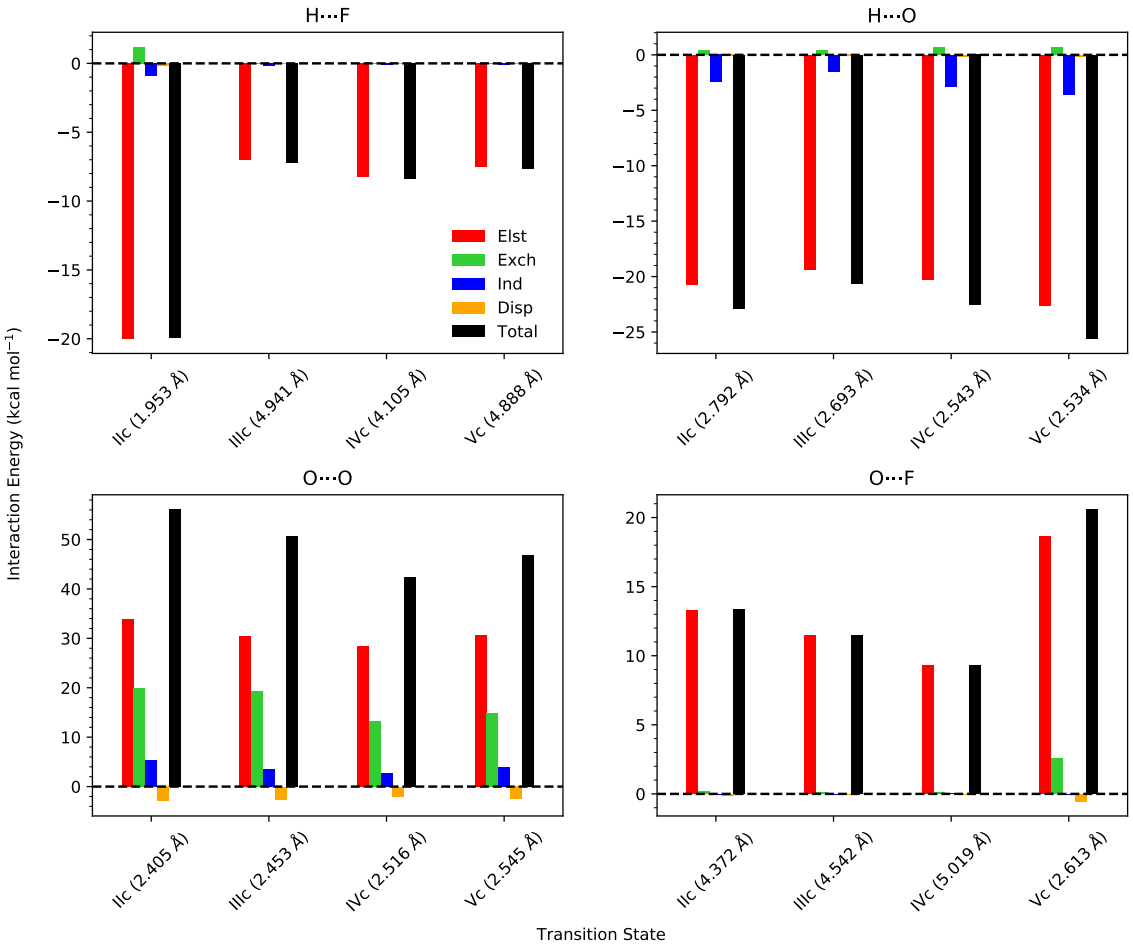


Figure 17: Exchange-scaled A-SAPT/jaDZ analysis of the interactions in transition states for Reaction C in terms of electrostatics, exchange-repulsion, induction, and dispersion. Each panel corresponds to an atom-atom interaction depicted in Figure 13.

TS's IV and V are clearly the most stabilized by this contact for reaction A, as expected by Hoveyda and co-workers on the basis of the contact distances (however, TS IV is slightly more stabilized despite a slightly longer distance, 2.59 vs 2.55 Å). V is the most stabilized TS in reactions B and C, despite having a longer contact distance than IV in reaction B (2.55 vs 2.54 Å). The contact in II becomes similarly stabilized as in IV for reactions B and C despite not having the chair conformation, and thus featuring a longer contact distance (about 2.80 Å in both reactions vs about 2.54 Å in IV). In reaction C, II is actually more stabilized by this interaction than IV. Across all three reactions, the hypothesis that V is stabilized relative to III by the H $\cdots$ O contact is supported by the A-SAPT analysis.

3.4.3 O ··· O

The aryloxy oxygen and carbonyl oxygen contact represents a close (2.39 – 2.57 Å), destabilizing interaction for all transition states of reactions A-C. However, it is expected to be more destabilizing in II and III due to a slightly closer contact distance (2.39 – 2.46 Å) in these TS's as compared to IV and V (2.52 – 2.57 Å). The O ··· O interaction is destabilizing by more than 40 kcal mol^{-1} across all examined TS's and is dominated by electrostatics (consistently around 30 kcal mol^{-1}). There are also considerable contributions from exchange-repulsion, which are markedly more destabilizing for TS's II and III (19 – 23 kcal mol^{-1}) than IV and V (13 – 16 kcal mol^{-1}) for reactions A-C. Induction/polarization and London-dispersion are relatively negligible components of all the O ··· O interactions, accounting for around 3 – 5 kcal mol^{-1} destabilization and around 2 – 3 kcal mol^{-1} stabilization, respectively.

As a result of the difference in exchange-repulsion, the total O ··· O interaction of IV and V (42 – 48 kcal mol^{-1}) is substantially less destabilizing than in II and III (51 – 57 kcal mol^{-1}) across all of the reactions. This difference in destabilization predicted by A-SAPT is consistent with Hoveyda and co-workers' hypothesis that the O ··· O interaction causes II to be energetically higher than IV for reactions A and B. Although the trends of this interaction remain the same for reaction C, the total DFT energy of IIc is either slightly lower than that of IVc, or very slightly higher, depending on the density functional employed,[40] which suggests that other factors must contribute to the energetic ordering for that reaction.

3.4.4 O ··· F

Repulsion between the aryloxy and fluorine in Va is hypothesized by Hoveyda and co-workers to make it less preferred than IVa, despite Va having presumably similar O ··· O and H ··· O interactions to IVa. This difference is expected due to the distance of the contact in V being around 2.7 Å, and over 4.4 Å for the other TS's, with IVa having the longest contact distance by a margin of $\sim$ 0.5 Å (similar distances are seen in reactions B and C). Indeed, in reactions A-C, A-SAPT reflects Hoveyda and co-workers' expectations:

TS IV minimizes the destabilizing effect of the O $\cdots$ F interaction (9 – 10 kcal mol^{-1}) and V maximizes the destabilization (19 – 21 kcal mol^{-1}).

Across all reactions, the electrostatic component is destabilizing (9 – 13 kcal mol^{-1} for II, III, and IV, and 17 – 19 kcal mol^{-1} for V). This is the only significant component of the interaction except in TS V, where the atoms are close enough for exchange-repulsion to become a minor contributor (2 – 3 kcal mol^{-1}). Thus, the A-SAPT results confirm the hypothesis that the O $\cdots$ F contact is destabilizing in V compared to IV or the other TS's.

Thus, the present A-SAPT analysis of the H $\cdots$ F, H $\cdots$ O, O $\cdots$ O, and O $\cdots$ F contacts are all consistent with the rationalizations of the DFT relative energies of the transition states reported by Hoveyda and co-workers.[40]

3.4.5 Steric Interactions in Reaction B

In reaction B, the methyl substituent of the fluoroketone substrate is replaced with a phenyl group. The DFT energies computed by Hoveyda and co-workers show that TS IV is now only 1.5–2.8 kcal mol^{-1} below II, compared to 2.6–3.3 kcal mol^{-1} for reaction A. They rationalize the decreased preference for TS IV by citing potentially destabilizing steric interactions between the bulk of the catalyst complex and the newly introduced phenyl group of the fluoroketone substrate. This hypothesis is supported by distortion of the carbonyl to phenyl O=C-C-C dihedral angle (32^{o} out of plane) in the substrate, which contributes to the removal of favorable π conjugation between the phenyl group and the carbonyl in TS IVb.

F-SAPT analysis shows that the interaction of the phenyl group of the substrate with the catalyst complex is slightly less favorable in IVb (-0.5 kcal mol^{-1}) than IIb (-0.6 kcal mol^{-1}) as seen in the top left panel of Figure 18; this is in accord with the expectations of Hoveyda and co-workers, but the energy difference is very small. Comparing the interaction between the phenyl group of the substrate and the phenyl group of the catalyst complex, shown in the top right panel of Figure 18, the exchange-repulsion is larger in IVb (5.5 kcal mol^{-1}) than in IIb (0.4 kcal mol^{-1}), as expected due to the much closer center of mass distance between the phenyl rings in IV (4.78 Å vs 6.73 Å). However, the structure of IVb places the

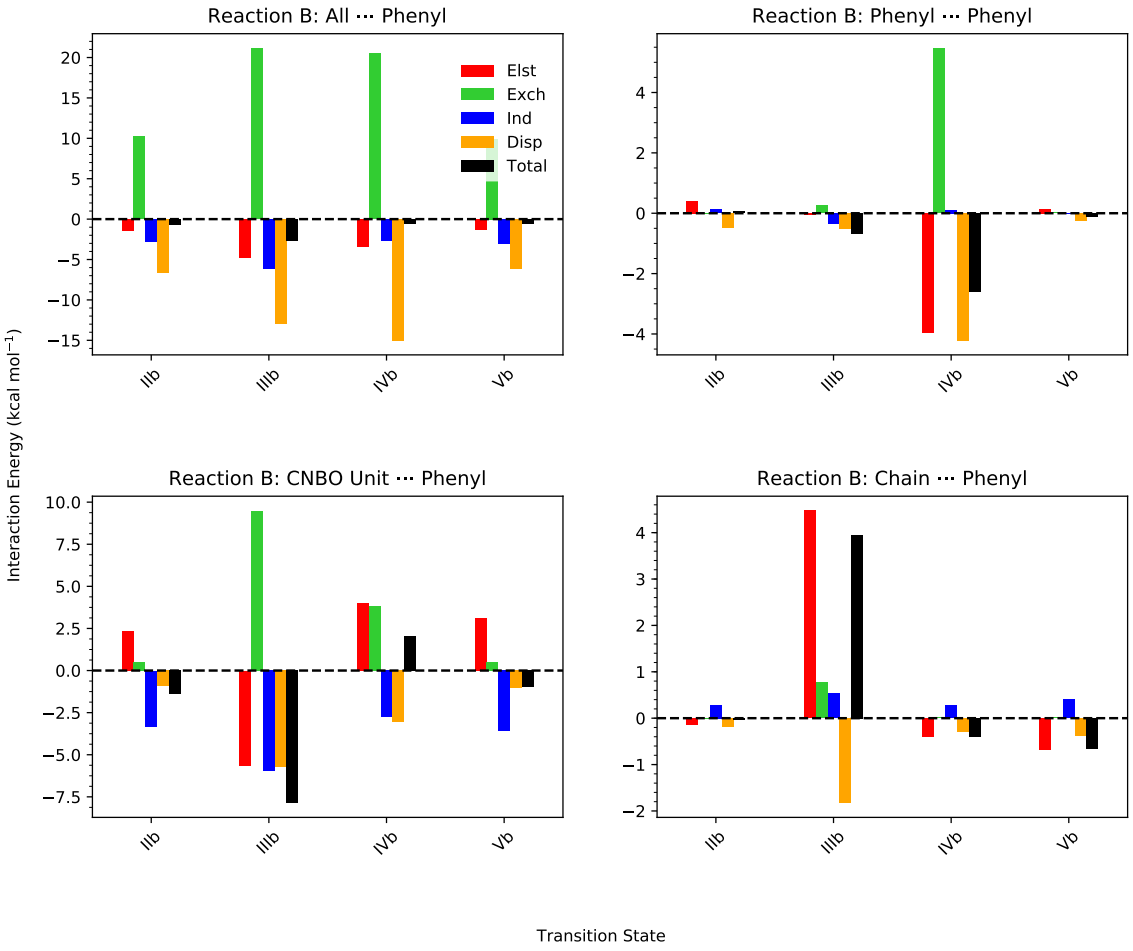


Figure 18: Exchange-scaled F-SAPT/jaDZ analysis of various interactions between the whole organoboron catalyst, phenyl, CNBO "chair", and the $(Me)_2$N-CO-CH-*t*-butyl "chain" of the organoboron catalyst with the phenyl of the substrate in transition states for Reaction B in terms of electrostatics, exchange-repulsion, induction, and dispersion.

closer phenyl rings in a favorable geometry for a π-π interaction, as shown in Figure 19 (the geometry is intermediate between an offset-stacked and a tilted T-shaped configuration), while the structure of IIb features a longer distance and poor relative orientation for a π-π interaction.

This favorable π-π geometry in IVb leads to stabilizing electrostatic contributions of 4.0 kcal mol^{-1} and London dispersion stabilization of 4.2 kcal mol^{-1} that overcome the higher exchange repulsion, leading to an overall phenyl-phenyl interaction energy of -2.6 kcal mol^{-1} in IVb. In IIb, the electrostatic component is destabilizing by 0.4 kcal mol^{-1},

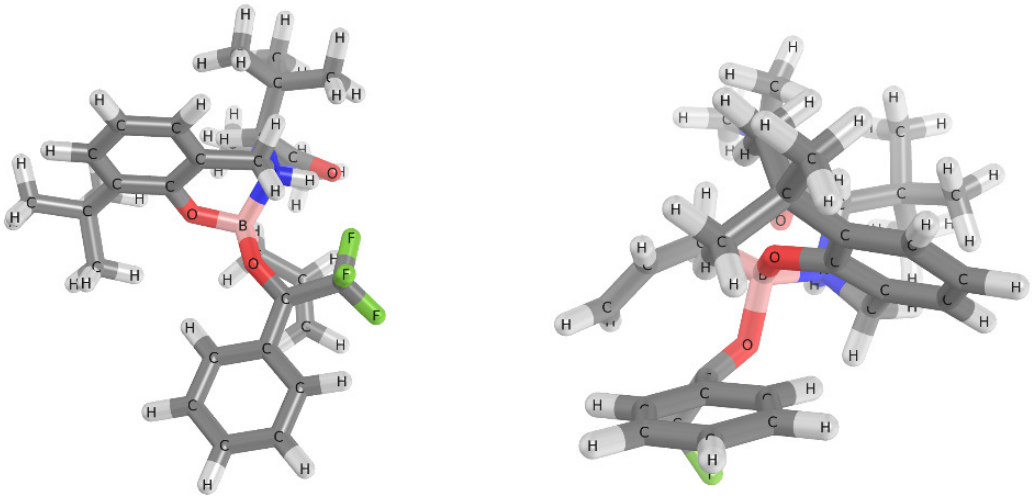


Figure 19: The π faces of IIb (left) do not align as favorably for π-π interactions as they align in IVb (right).

and London dispersion is stabilizing by 0.5 kcal mol^{-1} for the phenyl-phenyl interaction. This results in a total phenyl-phenyl interaction that is nearly zero (0.1 kcal mol^{-1}) in IIb. Thus, with regard to the phenyl-phenyl interaction, Hoveyda and co-workers are correct that the exchange-repulsion term is larger in IVb than IIb, but one might not have expected that a favorable π-π interaction will overcome this repulsion.

The interaction between a *t*-butyl substituent of the phenyl group on the catalyst and the phenyl of the substrate presents another possibility for increased steric repulsion in TS IVb relative to IIb. However, F-SAPT analysis shows this interaction in IVb is overall attractive (-0.8 kcal mol^{-1}) due to attractive electrostatic and dispersion terms overcoming a large, destabilizing exchange-repulsion term (4.2 kcal mol^{-1}). The same interaction in IIb is similarly stabilizing (-0.6 kcal mol^{-1}); however, individual components in this case are quite small because of the greater separation between the *t*-butyl and the phenyl group.

The phenyl group of the fluoroketone substrate also has significant interactions with the central CNBO "chair" of the catalyst in both IIb and IVb, as seen in the bottom left panel of Figure 18. In fact, among interactions involving the substrate phenyl, this interaction makes the largest difference towards destabilizing IV compared to II as one moves from Reaction A to reaction B (it shifts the relative interaction energies by 4.1 kcal mol^{-1}).

In IVb, there is a close contact (2.39 Å) between a hydrogen of the substrate phenyl and a hydrogen of the chair unit, leading to significant destabilizing exchange-repulsion and electrostatic terms. In IIb, the closest phenyl–CNBO contact is much more distant (3.27 Å), reducing the unfavorable interactions and yielding an overall 3.4 kcal mol^{-1} interaction energy preference for IIb over IVb. By contrast, in Reaction A, the CNBO chair / substrate methyl interaction is much more similar in IIa and IVa, and IVa is actually slightly preferred by this interaction (by 0.7 kcal mol^{-1}).

The most significant remaining interaction involving the substrate phenyl group is a rather unfavorable electrostatic interaction between the phenyl group and the $(Me)_2$N-CO-CH-*t*-butyl "chain" of the catalyst, which further destabilizes the relative interaction energy of IIIb compared to what it was in IIIa (by about 4 kcal mol^{-1}, see bottom right panel of Figure 18).

Finally, although it does not directly involve the substrate R group (methyl or phenyl), we note that the catalyst chair / substrate carbonyl interaction is significantly less stabilizing in IVb relative to IIb (-3.9 kcal mol^{-1}) than it was in IVa relative to IIa (-8.3 kcal mol^{-1}, see Ancillary Material). This interaction includes the incipient bond formation between the catalyst boron atom and the carbonyl oxygen of the substrate, which involves large energy components. IVa and IVb feature similar contact distances (1.59 and 1.58 Å, respectively), and IVb is 4.0 kcal mol^{-1} more stabilized by this contact than IVa (although the difference is only about 9% for this large interaction energy). By contrast, the geometry of IIb allows for a much closer contact (1.54 Å) than is possible in IIa (1.62 Å), meaning that the improvement in the interaction energy of IIb vs. IIa is much larger (8.4 kcal mol^{-1}). Thus, differences in the transition state geometries mean that the catalyst chair / substrate carbonyl interaction also contributes to the decreased preference of TS IV vs TS II as one moves from Reaction A to Reaction B.

To summarize this section, we have investigated a hypothesis of Hoveyda and co-workers[40] to explain why the DFT energy gap between TS's IV and II is reduced by 0.5–1.1 kcal mol^{-1} in Reaction B compared to Reaction A. Those researchers suggested that the decreased preference for IV is due to larger steric repulsions between the catalyst

and the substrate, due to the larger R group in Reaction B (phenyl) compared to reaction A (methyl). The expected steric clash between the phenyl of the substrate and the phenyl of the catalyst is, contrary to prior expectations, *stabilizing* in TS IVb due to a favorable π-π interaction. Likewise, the substrate phenyl has favorable interactions with the *t*-butyl group off the catalyst phenyl (which are similar in magnitude for TS's IVb and IIb). The biggest direct contributor to the destabilization of IVb vs IIb involving the substrate phenyl is the interaction of the substrate phenyl with the central CNBO chair of the catalyst, but this contribution is canceled by the previously mentioned substrate phenyl interactions that are more favorable for IVb than IIb, so that the net phenyl $\cdots$ catalyst interactions are approximately equal for IVb and IIb. The most important change in the relative interaction energies of IV and II as one moves from Reaction A to Reaction B seems to be a change in the interaction between the catalyst CNBO chair and the substrate carbonyl, where IV shows a reduced preference over II in Reaction B compared to Reaction A, which is caused by a more favorable bond formation geometry in IIb compared to IIa.

Finally, we note that in terms of just interaction energies, IVb is 2.4 kcal mol^{-1} *less* stable than IIb, whereas IVa is 4.5 kcal mol^{-1} more stable than IIa (see Figure 14). The fact that IVb remains below IIb according to Hoveyda's DFT computations is attributed to differences in reactant distortion energies. Indeed, at the MP2/aTZ level of theory, we compute that the distortion energy of IIb is nearly 6 kcal/mol greater than that of IVb (see Figure 41 of the Ancillary Material).

3.4.6 Steric Interactions in Reaction C

In reaction C, an *ortho*-methyl substituent has been added to the phenyl group of the substrate. Hoveyda and co-workers found that this addition reduced the enantiomeric ratio to 83:17, compared to 96:4 for Reaction B.[40] They hypothesized that the additional steric bulk from the methyl group led to more repulsion between the substrate and the catalyst's phenol moiety, thus destabilizing the transition states (IIc and IVc) that lead to the preferred enantiomer; the other two transition states (IIIc and Vc) orient the R group of the substrate away from the phenol group of the catalyst, and are not affected by the additional steric

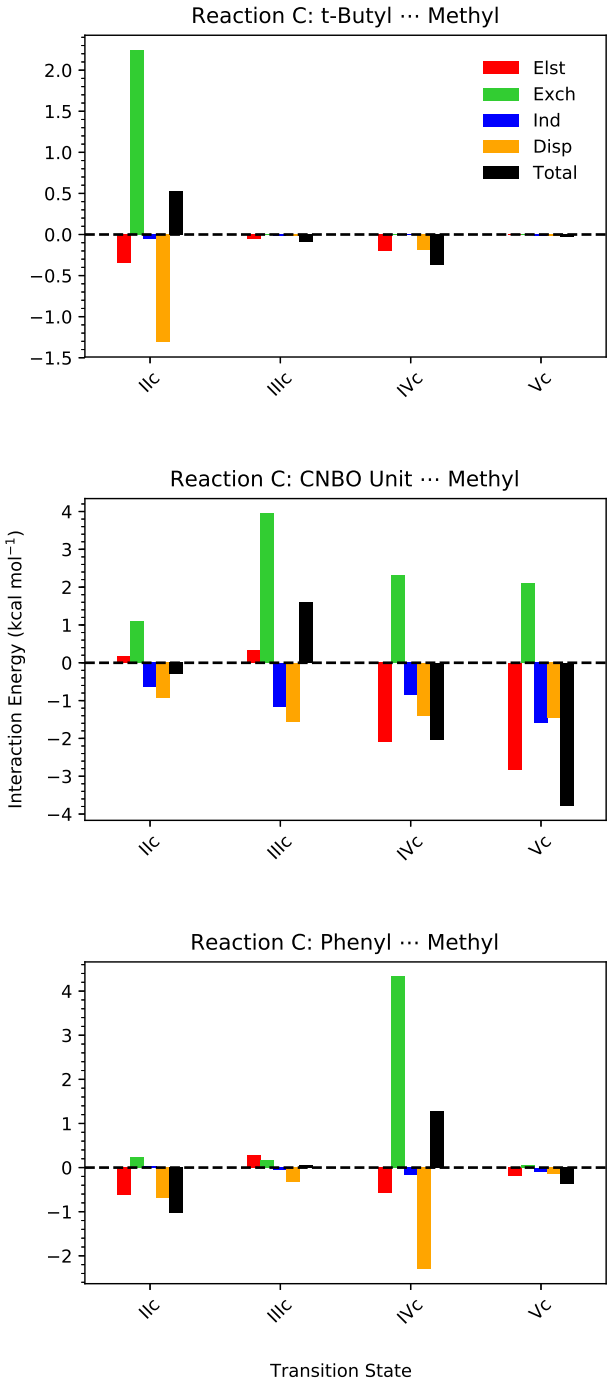


Figure 20: Exchange-scaled F-SAPT/jaDZ analysis of various interactions between the *t*-butyl, phenyl, and CNBO "chair" of the organoboron catalyst with the *ortho*-methyl group of the substrate in transition states for Reaction C in terms of electrostatics, exchange-repulsion, induction, and dispersion.

bulk of the methyl group. The DFT results tend to support this hypothesis. Transition state Vc is now only 1.4 or 1.8 kcal mol^{-1} (M06-2X or ωB97X-D) above the most stable transition state (either IIc or IVc, depending on the functional); in Reaction B, the gap is somewhat larger, 1.7 or 3.2 kcal mol^{-1}, respectively. This makes the non-preferred reaction pathway less unfavorable. (Transition state III remains higher in energy by several kcal mol^{-1}).

Moreover, the notion that this decreased preference for IIc and IVc relative to Vc is due to steric clashes between the R group of the substrate and the phenol moiety of the catalyst is supported by the transition state geometries. TS IVc demonstrates an even larger distortion away from planar conjugation for the carbonyl and phenyl groups of the substrate (O=C-C-C angle of $\sim 55^{o}$) than was seen for Reaction B ($\sim 32^{o}$). In TS IIc, there is a close 2.33 Å contact between the *ortho*-methyl substituent and the *t*-butyl group of the catalyst.

SAPT analysis supports the idea that TS's II and IV are destabilized relative to V as the *ortho*-methyl substituent is added to the substrate phenyl ring. As shown in Figure 14, for Reaction B, TS V has about the same amount of interaction energy stabilization as IV, while II has an additional 2-3 kcal mol^{-1} of stabilization. For Reaction C, the gap between II and V is reduced to only $\sim$ 1 kcal mol^{-1} (II remains more stable), while IV's interaction energy is now 7 kcal mol^{-1} less favorable than II's. (The fact that IVc's DFT energy is very similar to those of IIc and Vc, despite IVc having a much less stabilizing interaction energy, is due to its having a lower deformation energy than IIc and Vc; MP2/aTZ computations show that IIc and Vc have deformation energies $\sim$ 8 kcal mol^{-1} higher than IVc, see Figure 41 of the Ancillary Material).

F-SAPT analysis provides more detailed information on the origins of these differences in non-covalent interaction energies. Here, we examine interactions between the *ortho*-methyl portion of the substrate vs the *t*-butyl, phenyl, and CNBO "chair" portions of the catalyst. The results are presented in Figure 20.

For the interaction of the *t*-butyl group off of the phenyl of the organoboron catalyst with the *ortho*-methyl of the substrate, IIc is expected to have the strongest destabilizing

interaction due to it having the closest contact distance (2.33Å). Indeed, as seen in the top left panel of Figure 20, IIc has the most destabilizing interaction according to F-SAPT. However, the substantial 2.2 kcal mol^{-1} exchange repulsion for this contact is mostly canceled by a -1.3 kcal mol^{-1} dispersion interaction; the overall contact including electrostatic terms also is repulsive by only 0.5 kcal mol^{-1}. This same interaction is negligible in IIIc and Vc, and is weakly stabilizing (-0.4 kcal mol^{-1}) in IVc.

A close contact between the phenyl group of the organoboron catalyst and the *ortho* methyl of the substrate is expected to result in steric clashing for TS IVc (C $\cdots$ H distance of 2.51 Å). According to F-SAPT analysis (bottom panel, Figure 20), the interaction is most destabilizing in IVc (1.3 kcal mol^{-1}). IIc's interaction is stabilizing by an almost equal magnitude (1.0 kcal mol^{-1}).

Though not explicitly mentioned by Hoveyda and co-workers, the CNBO unit of the organoboron catalyst appears to sterically clash with the *ortho*-methyl group in IIIc, contributing to the relative interaction energy of III being even higher in reaction C than it was in reaction B (see Figure 14). F-SAPT analysis confirms that IIIc is destabilized by this contact (by 1.6 kcal mol^{-1}), although IVc and Vc are substantially stabilized (by 2.1 kcal mol^{-1} and 3.8 kcal mol^{-1}, respectively, middle panel of Figure 20). As anticipated, exchange-repulsion is the most destabilizing component of the interaction IIIc (4.0 kcal mol^{-1}), and is large enough to overcome the stabilizing dispersion and induction terms (which sum to 2.7 kcal mol^{-1}). IVc and Vc have significant electrostatic stabilization for this contact, with additional stabilizing dispersion and induction contributions. The significant stabilization of Vc by this contact contributes to the nearly equal interaction energies of IIc and Vc in Figure 14.

Overall, F-SAPT analysis supports the idea that steric clashes involving the methyl substituent on the substrate phenyl lead to a less favorable interaction energy of IVc relative to Vc. The *t*-butyl $\cdots$ *ortho*-methyl contact also destabilizes IIc vs Vc, but by a modest amount that is approximately canceled by a more favorable phenyl $\cdots$ methyl contact in IIc vs Vc. Perhaps surprisingly, favorable contacts between the methyl substituent and the CNBO "chair" of the catalyst are also substantially stabilizing for Vc vs the other transition

states.

3.5 Conclusions

Work by Hoveyda and co-workers revealed that the oraganoboron catalyzed allyl addition to a ketone substrate results in an enantioselective reaction upon adding fluorines to the ketone substrate.[40] When acetophenone is the substrate, the preferred enantiomer is produced with low efficiency (32:68 e.r.); however, the fluorinated ketone (2,2,2-trifluoroacetophenone) produces the desired enantiomer through transition states II and IV in an enantioselective reaction (96:4 e.r.). Hoveyda and co-workers posit that this enhanced selectivity is due to the presence of a stabilizing electrostatic interaction between a fluorine of the ketone substrate and a nearby ammonium proton in the transition states leading to the dominant product. Additional non-covalent interactions are invoked to explain the relative energies of the transition states as obtained by density functional theory. The current work has analyzed these claims by direct quantification of non-covalent interactions using the atomic and functional group partitions of symmetry-adapted perturbation theory (A-SAPT and F-SAPT).

As hypothesized by Hoveyda and co-workers, A-SAPT confirms that there is a large stabilization of the transition states due to the non-covalent interaction between the substrate fluorine atom and the ammonium proton of the catalyst, and that this stabilization is much greater for one of the transition states (II) leading to the preferred enantiomer. This stabilization is almost entirely electrostatic in nature. Other atom-atom interactions proposed to influence the enantioselectivity of the reaction by Hoveyda and co-workers (carbonyl oxygen of the substrate interacting with the ammonium proton of the organoboron catalyst, the carbonyl oxygen interacting with a nearby oxygen in the CNBO unit of the organoboron catalyst, and the oxygen of the CNBO unit of the organoboron catalyst interacting with the nearest fluorine of the fluoroketone substrate) are also confirmed by our A-SAPT analysis.

We examined three reactions studied by Hoveyda and co-workers; in Reaction A, the R group of the trifluoro-methyl ketone substrate is a methyl group, whereas in Reactions B, and C, the methyl of the substrate is replaced by a phenyl group or an *ortho*-phenyl

group, respectively. For reactions B and C, the bulkier R-group was expected to destabilize the transition states through steric clashing with the organoboron catalyst. However, F-SAPT analysis shows that this interaction is actually a stabilizing π-π interaction, and it contributes to the overall preference for TS IV.

Overall, this work has directly confirmed a majority of the non-covalent interactions that were thought to play a role in the enantioselectivity of the organoboron catalyzed allyl addition to a fluoroketone through the application of A-SAPT and F-SAPT. F-SAPT analysis suggests that some interactions that may be seen as sterically repulsive according to chemical intuition are actually stabilizing when electrostatic and dispersion interactions are quantified. The results of this study suggest that these SAPT partitions can be used to directly assess the relative strengths on non-covalent interactions in future organocatalysis research.

CHAPTER IV

SUBSTITUTION OF LOCAL MP2 IN B2PLYP

4.1 Summary

Double hybrid density functional theory (DH-DFT) combines hybrid density functional theory with a more rigorous description of long range dispersion from Møller-Plesset second-order perturbation theory (MP2). Unfortunately, the computational cost of MP2 is fairly steep, scaling as $O(N^5)$ where N is proportional to molecular size. However, there are lower scaling yet accurate methods to account for electron correlation, such as local MP2 (LMP2). This work serves as a systematic investigation of the effect of substituting LMP2 for MP2 in double hybrid DFT. Statistics from S22 database computations in the cc-pVDZ and cc-pVTZ basis sets show that the average difference between the canonical B2PLYP and local B2PLYP (LB2PLYP) results is 0.189 kcal mol^{-1} and 0.529 kcal mol^{-1} on average with and without counterpoise correcting, respectively.

4.2 Introduction

Fifty years after its introduction, density functional theory (DFT) has become an invaluable tool for applications throughout chemistry and physics.[37] Pure DFT has achieved this popularity because of its ability to provide wavefunction theory (WFT) quality results for a lower cost than the cheapest WFT method, Hartree-Fock.[12] Despite its great utility, pure DFT has its drawbacks, including inadequate descriptions of exchange and correlation. To correct for this, Hartree-Fock exchange has been added to the exchange-correlation functional in many approaches, termed hybrid DFT functionals,[6,7] which makes this class of methods approximately the same cost as Hartree-Fock, but with higher accuracy.

Double hybrid DFT was first introduced by Grimme in 2006 and was intended to capture electron correlation more accurately than previous hybrid DFT functionals by adding a term from second-order perturbation theory to the exchange-correlation functional.[23] Unfortunately this addition means that the cost is now dominated by the high order scaling

$O(N^5)$ of MP2. To reduce this additional computational expense, one may consider using local correlation approximations for the MP2 component.

The local approximation was largely pioneered by Saebø and Pulay[54] and has been further developed into its modern form by other groups, such as Werner et al.[26,62] Local correlation versions of MP2 and coupled-cluster theory have been implemented and applied to study large systems,[62,81] including the protein crambin.[60] Asymptotically linear scaling has been achieved by the use of local methods in these cases. Additionally, the local approximation has been shown to reduce basis set superposition error (BSSE), which opens up the possibility of avoiding the expensive counterpoise correction for van der Waals dimers.[26]

The current study specifically aims to compare B2PLYP with local MP2 (LB2PLYP) to B2PLYP on the basis of interaction energies. Other work has shown that B2PLYP is highly accurate for predicting many molecular quantities of interest (vibrational properties, interaction energies, *etc.*)[8,12] , which is promising for our new approach. Recently, Sharkas *et al.* have computed lattice energies for molecular crystals with DH-DFT using periodic LMP2 in conjunction with B2PLYP and other DH-DFT functionals,[69] and they found it to be the worst performing DH-DFT functional tested. However, periodic LMP2 was found to generally perform better than all DFT functionals tested. There was no comparison to canonical B2PLYP reported in their study, as we are presenting here.

4.3 Theoretical Methods

4.3.1 DH-DFT Methods

The DH-DFT method B2PLYP[23] is named for its components: B for B88 exchange functional,[5] 2 for second-order perturbation theory and the number of parameters used, and PLYP for perturbative correlation in addition to the LYP functional.[41] The exchange-correlation functional for B2PLYP can be written as a linear combination:[23]

$$E_{xc} = (1 - a_x)E_x^{GGA} + a_x E_x^{HF} + bE_c^{GGA} + cE_c^{PT2}, \tag{48}$$

where $a_x = 0.53$, $b = 0.73$ and $c = 0.27$. Here, the E_x^{GGA} and E_x^{HF} terms account for the exchange energy captured using Kohn-Sham (KS) orbitals with the Generalized Gradient

Approximation (GGA) and Hartree-Fock (HF), respectively. Similarly, the E_c^{GGA} and E_c^{PT2} terms account for the correlation energy captured using GGAs and second order Møller-Plesset perturbation theory (MP2), respectively. The value chosen for the HF exchange mixing parameter, a_x, indicates that just over half of the exchange is being accounted for by a HF description and the rest of exchange is coming from the GGA approximation. As prescribed by Grimme, $b = 1 - c$, where it can be seen a majority (73%) of the electron correlation is being accounted with the GGA approximation. This much smaller percentage (27%) accounted for by MP2 means the ultimate result of the B2PLYP functional will not be greatly affected by the MP2 contribution if it is not large. In the present work, we consider how B2PLYP results are affected if we invoke the local correlation approximation for E_c^{PT2}, yielding what we will designate as local B2PLYP, or LB2PLYP.

4.3.2 Local Correlation Methods

A local MP2 procedure inspired by Schütz *et al.*[62] was used to compute the second-order perturbation term in the above B2PLYP ansatz for this work. For this method, the orbitals are localized using a localization scheme, such as the Boys or Pipek-Mezey methods.[10,53] It is known that the Pipek-Mezey scheme preserves $\sigma-\pi$ symmetry but otherwise yields similar results to Boys.[26] Boys localization was chosen for this work. The local molecular orbitals (LMOs) span the occupied space in LMP2, and a set of non-orthogonal projected atomic orbitals (PAOs) span the virtual space. LMOs are only allowed to excite into a spatially close subset of PAOs, or orbital domain, that is assigned according to a procedure developed by Boughton and Pulay.[9] Finally, the electron pairs are treated differently depending on the separation distance of the LMOs. For instance, pairs within 1 Bohr of each other are considered *strong pairs* and are treated fully within the local MP2 methodology. The pairs between 1 and 8 Bohr away are called *weak pairs* and are treated with full LMP2 as well. If pairs are further apart, 8 to 15 Bohr (*distant pairs*), a multipole approximation may be used, and, if the pairs are more than 15 Bohr apart, they may be neglected entirely without much loss of energy.[26] Please see section 1.2.2 for more details on how this procedure works.

4.3.3 Computations

In order to gauge the performance of LB2PLYP, the interaction energies for Hobza's S22 test set of van der Waals dimers[35] was considered. This set contains 8 hydrogen bonded dimers, 7 dispersion dominated dimers, and 7 dimers of mixed character. Benchmark energies for the S22 test set were taken from more recent data in the S22B revision.[45] All LB2PLYP computations were performed with the open source Psi4 quantum chemistry package[74] under the frozen core approximation and using a density fitted SCF procedure. Grimme's -D3 dispersion correction[24] was applied to all computations in this work because it has been recommended to generally provide the best results for general non-covalent interactions by Burns *et al.*[12] The Boys localization procedure was used throughout to obtain local orbitals from the Kohn-Sham orbitals. The LMP2 code, as recommended by Werner *et al*,[26] employed a Boughton-Pulay threshold of 0.02 and fully treated all pairs within 15 Bohr using the traditional MP2 equations. Additionally, the density fitting approximation was applied for the two-electron integrals required in LMP2. Combinations of overlap functions corresponding to eigenvalues smaller than 1.0e-6 were discarded from the overlap matrix when orthogonalizing the PAOs to avoid numerical problems. All computations either made use of the cc-pVDZ (DZ) or cc-pVTZ (TZ) basis sets and associated JKFIT[76] and RI[79] fitting basis sets for the SCF and LMP2 density fitting procedures, respectively.

4.4 Results and Discussion

We first consider a comparison of the B2PLYP and LB2PLYP results for the interaction energies of the van der Waals dimers of the S22 test set. Figures 21 and 22 show these results in the DZ and TZ basis sets, respectively, and the reference CCSD(T) complete basis set limit [CCSD(T)/CBS] values. Both counterpoise corrected (CP) and uncorrected (unCP) results are presented. Figure 23 plots the mean difference in interaction energies between LB2PLYP and B2PLYP for S22 and its previously mentioned subsets. Last, Figure 24 provides a breakdown comparison of the mean absolute error statistics (MAEs) for LB2PLYP and B2PLYP vs. the reference CCSD(T)/CBS values over S22 and the same

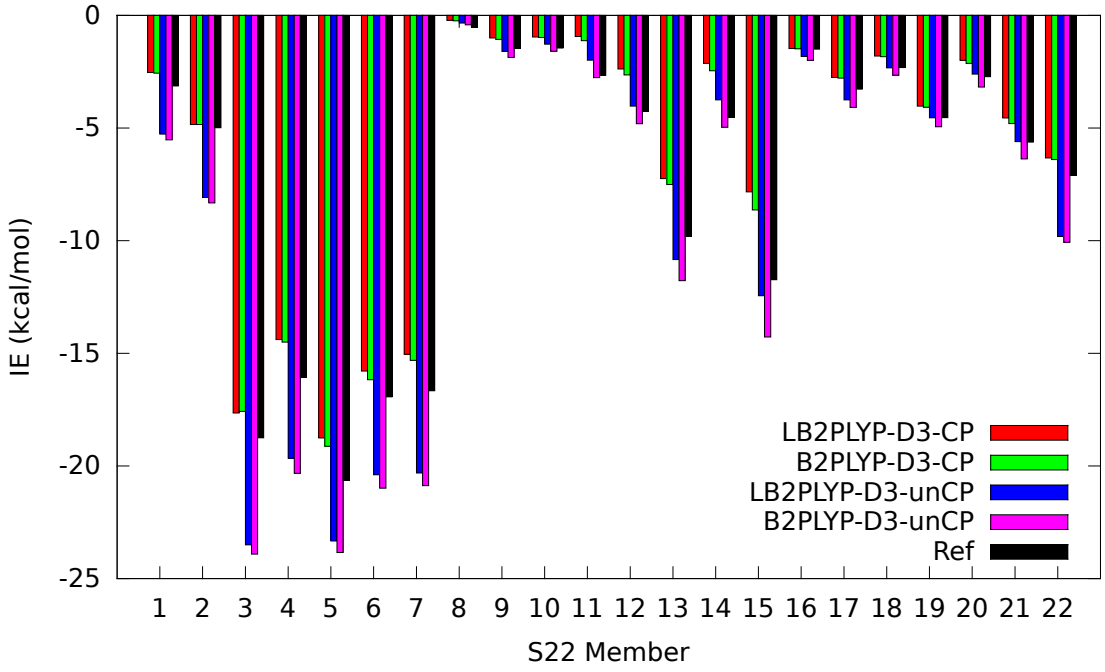


Figure 21: The results for interaction energy calculations on the S22 database using the B2PLYP and LB2PLYP with and without counterpoise correcting in the cc-pVDZ (DZ) basis set. Also, the reference values from the S22B revised benchmarks are plotted. The x-axis corresponds to the S22 dimer number, and the y-axis represents the interaction energy (IE) of the dimer in kcal mol^{-1}.

subsets. Additionally, all data is presented in the form of Tables 2 – 5 for future reference.

4.4.0.1 LB2PLYP vs B2PLYP

Our primary goal is to show that LB2PLYP provides similar results to B2PLYP for a diverse group of non-covalent interactions. The maximum difference between LB2PLYP and B2PLYP in both basis sets was the unCP computation for adenine-thymine dimer in a parallel geometry (S22 #15): 1.830 and 1.721 kcal mol^{-1} for the DZ and TZ basis set, respectively, out of a benchmark CCSD(T)/CBS interaction energy of 11.730 kcal mol^{-1} (i.e. around 15%). The mean average difference between IEs from LB2PLYP and B2PLYP for the S22 test set are shown in Figure 23. It can immediately be seen that, on average, the difference in IE between the local and canonical result is always under ∼0.65 kcal mol^{-1} regardless of the choice of basis set and counterpoise correction. For all of S22, improving the basis from DZ to TZ translates to CP results becoming ∼0.03 kcal mol^{-1} farther apart

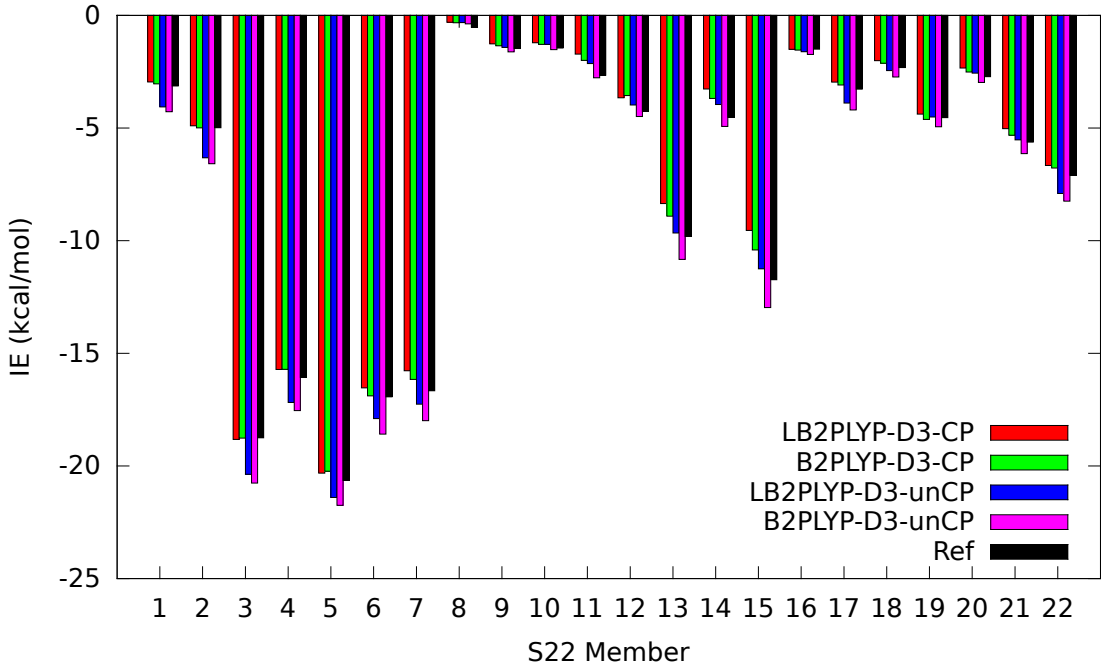


Figure 22: The results for interaction energy calculations on the S22 database using the B2PLYP and LB2PLYP with and without counterpoise correcting in the cc-pVTZ (TZ) basis set. Also, the reference values from the S22B revised benchmarks are plotted. The x-axis corresponds to the S22 dimer number, and the y-axis represents the interaction energy (IE) of the dimer in kcal mol^{-1}.

and unCP results becoming ∼0.06 kcal mol^{-1} closer.

Increasing the basis set does not have much of an effect on how the B2PLYP variants compare across the subsets of S22. In all subsets, an increase in basis set size causes the unCP results to agree more closely on average. Notably, the unCP results for DD complexes are on average ∼0.1 kcal mol^{-1} closer in the larger cc-pVTZ basis. The interaction energies computed by the two versions of B2PLYP for HB complexes become closer by ∼0.04 kcal mol^{-1}, and the MX complexes stayed virtually the same, becoming more similar by ∼0.01 kcal mol^{-1} in the larger basis set. The CP results only become closer, by ∼0.02 kcal mol^{-1}, in for the HB subset. In the DD and MX subsets, the results become farther apart by ∼0.02 and 0.1 kcal mol^{-1}, respectively.

The counterpoise correction has a similar effect for all of the subsets: the CP results are in better agreement than unCP. For CP results, HB and DD systems have an average

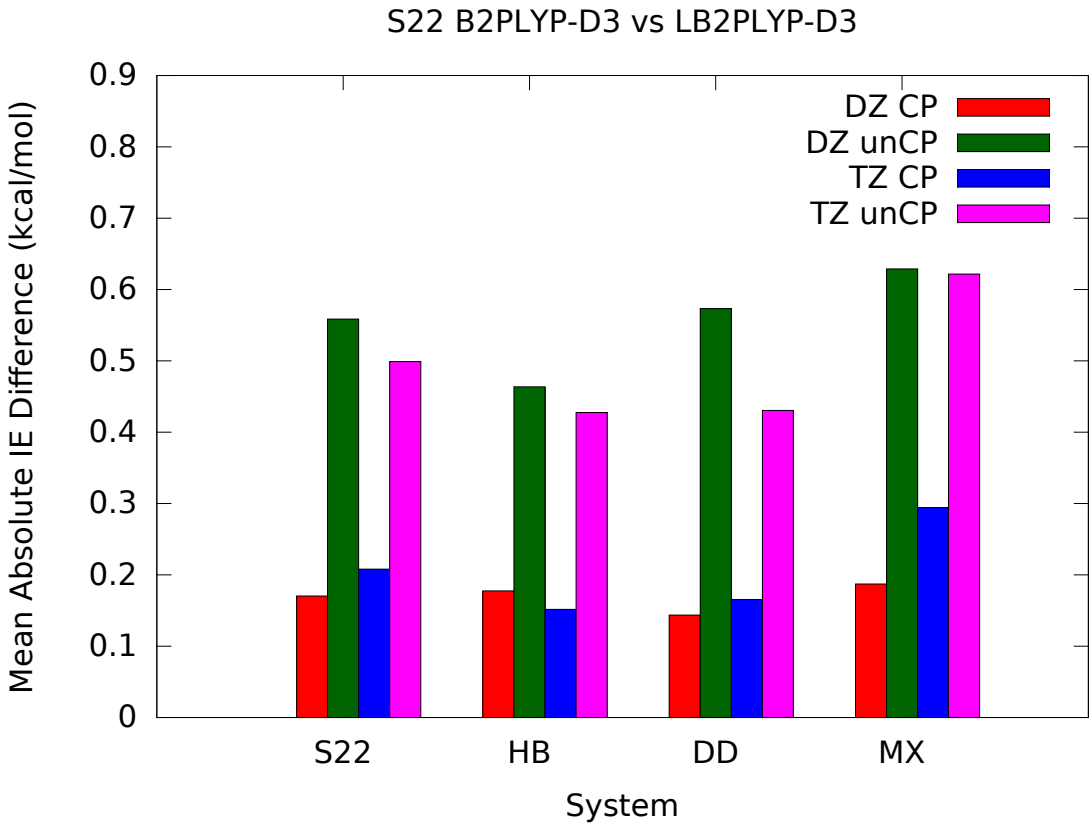


Figure 23: Mean absolute interaction energy differences between B2PLYP-D3 and LB2PLYP-D3 are plotted for S22 and the hydrogen bonding (HB), dispersion dominated (DD), and mixed interaction (MX) subsets of S22 computed in the cc-pVDZ (DZ) and cc-pVTZ (TZ) basis sets. Both results with (CP) and without the counterpoise correction are presented in kcal mol^{-1}.

agreement of $\sim$0.15 kcal mol^{-1} no matter the basis set, and the MX results have $\sim$0.2 or $\sim$0.3 kcal mol^{-1} agreement in the DZ and TZ basis sets respectively. For unCP computations, the HB subset had the closest agreement on average at $\sim$0.45 kcal mol^{-1} in either basis set. The DD subset had the worst agreement on average at $\sim$0.62 kcal mol^{-1} in either basis set. The DD agreement was in between the other subsets on average at $\sim$0.57 kcal mol^{-1} in the DZ basis and $\sim$0.43 kcal mol^{-1} in the TZ basis set.

4.4.0.2 LB2PLYP vs CCSD(T)/CBS

Judging from the MAEs in Fig 24, one can see that the errors in all approaches test decrease with an increase in basis set size. In the DZ basis, the MAE for all of S22 ranges from ∼1.0-1.5 kcal mol^{-1} for all approaches. The MAEs decrease to range from ∼0.3-0.8 kcal mol^{-1} in the TZ basis. Additionally, both B2PLYP variants yield relatively accurate results whether counterpoise correcting or not. Considering the S22 test set as a whole, the CP results for both B2PLYP variants give slightly lower MAEs than the unCP results. Notably, however, the CP results are more accurate for hydrogen bonding dominated interactions while the unCP results are better for dispersion dominated interactions.

Table 2: Computed interaction energies for B2PLYP variants in the cc-pVTZ basis and benchmark values are listed.

S22	Ref	LB2PLYP-D3-CP	LB2PLYP-D3-unCP	B2PLYP-D3-CP	B2PLYP-D3-unCP
1	-3.133	-2.536	-5.271	-2.567	-5.528
2	-4.989	-4.841	-8.082	-4.847	-8.326
3	-18.753	-17.649	-23.502	-17.580	-23.916
4	-16.062	-14.388	-19.661	-14.502	-20.325
5	-20.641	-18.758	-23.329	-19.125	-23.844
6	-16.934	-15.789	-20.392	-16.175	-20.986
7	-16.660	-15.045	-20.312	-15.315	-20.870
8	-0.527	-0.224	-0.339	-0.241	-0.416
9	-1.472	-0.999	-1.596	-1.063	-1.870
10	-1.448	-0.961	-1.280	-0.981	-1.591
11	-2.654	-0.940	-1.985	-1.121	-2.763
12	-4.255	-2.380	-4.029	-2.639	-4.808
13	-9.805	-7.244	-10.837	-7.511	-11.779
14	-4.524	-2.130	-3.747	-2.458	-4.961
15	-11.730	-7.834	-12.445	-8.637	-14.275
16	-1.496	-1.467	-1.825	-1.484	-1.999
17	-3.275	-2.759	-3.750	-2.781	-4.085
18	-2.312	-1.806	-2.329	-1.831	-2.657
19	-4.541	-4.029	-4.546	-4.078	-4.942
20	-2.717	-1.993	-2.603	-2.129	-3.185
21	-5.627	-4.556	-5.601	-4.798	-6.371
22	-7.097	-6.327	-9.814	-6.399	-10.071

Table 3: Computed interaction energies for B2PLYP variants in the cc-pVTZ basis and benchmark values are listed.

S22	Ref	LB2PLYP-D3-CP	LB2PLYP-D3-unCP	B2PLYP-D3-CP	B2PLYP-D3-unCP
1	-3.133	-2.952	-4.061	-3.038	-4.275
2	-4.989	-4.902	-6.322	-4.991	-6.583
3	-18.753	-18.824	-20.372	-18.765	-20.760
4	-16.062	-15.716	-17.180	-15.711	-17.544
5	-20.641	-20.321	-21.399	-20.235	-21.745
6	-16.934	-16.536	-17.900	-16.886	-18.584
7	-16.660	-15.774	-17.256	-16.161	-17.990
8	-0.527	-0.304	-0.323	-0.330	-0.379
9	-1.472	-1.263	-1.423	-1.350	-1.620
10	-1.448	-1.217	-1.288	-1.288	-1.521
11	-2.654	-1.722	-2.139	-1.999	-2.772
12	-4.255	-3.656	-3.976	-3.559	-4.488
13	-9.805	-8.350	-9.658	-8.914	-10.833
14	-4.524	-3.267	-3.959	-3.689	-4.932
15	-11.730	-9.548	-11.248	-10.414	-12.969
16	-1.496	-1.503	-1.611	-1.544	-1.731
17	-3.275	-2.960	-3.895	-3.087	-4.198
18	-2.312	-2.008	-2.454	-2.124	-2.735
19	-4.541	-4.380	-4.513	-4.615	-4.941
20	-2.717	-2.336	-2.562	-2.514	-2.974
21	-5.627	-5.032	-5.526	-5.320	-6.133
22	-7.097	-6.659	-7.908	-6.777	-8.244

Table 4: Mean absolute errors (MAEs) relative to benchmark values for S22 interaction energy computations in the cc-pVDZ basis are listed.

System	LB2PLYP-D3-CP	LB2PLYP-D3-unCP	B2PLYP-D3-CP	B2PLYP-D3-unCP
S22	1.182	1.407	1.018	1.779
HB	1.166	3.340	1.009	3.803
DD	1.139	0.323	0.995	0.317
MX	1.233	0.665	1.046	1.287

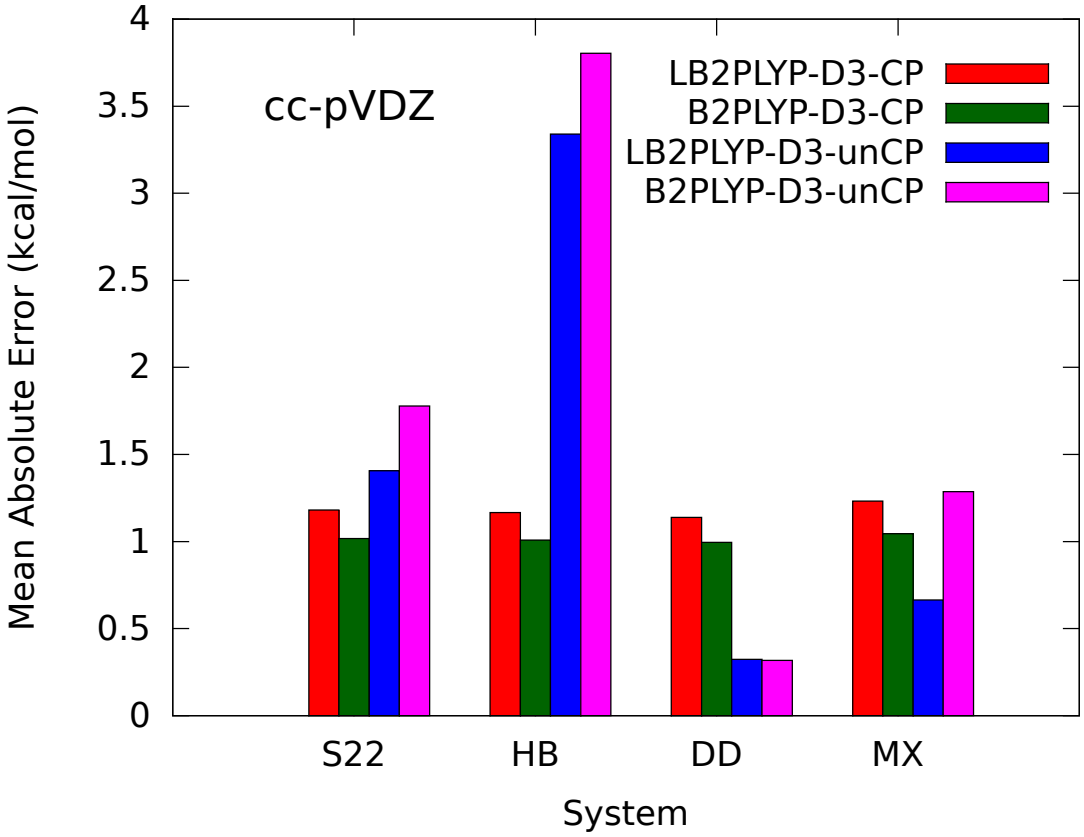
Table 5: Mean absolute errors (MAEs) relative to benchmark values for S22 interaction energy computations in the cc-pVTZ basis are listed.

System	LB2PLYP-D3-CP	LB2PLYP-D3-unCP	B2PLYP-D3-CP	B2PLYP-D3-unCP
S22	0.526	0.531	0.346	0.800
HB	0.327	1.045	0.202	1.473
DD	0.548	0.275	0.410	0.198
MX	0.682	0.306	0.417	0.738

4.5 Conclusions

This study presents a performance comparison for a local version of the popular DH-DFT functional, B2PLYP, and its canonical form. DH-DFT functionals add a second-order perturbation theory term to account for electron correlation, which typically entails MP2. This addition dominates the previous cost of DFT making this approach unfeasible for large systems. By making the local approximation, larger systems should be accessible with the same high accuracy of canonical B2PLYP. The benchmark was performed using S22, which revealed a similar performance with the differences between the local and canonical versions of B2PLYP being $\sim$0.65 kcal mol^{-1} different on average regardless of basis set or

counterpoise correction choices. Our results show that LB2PLYP can replicate the quality of B2PLYP through the use of a local correlation method.



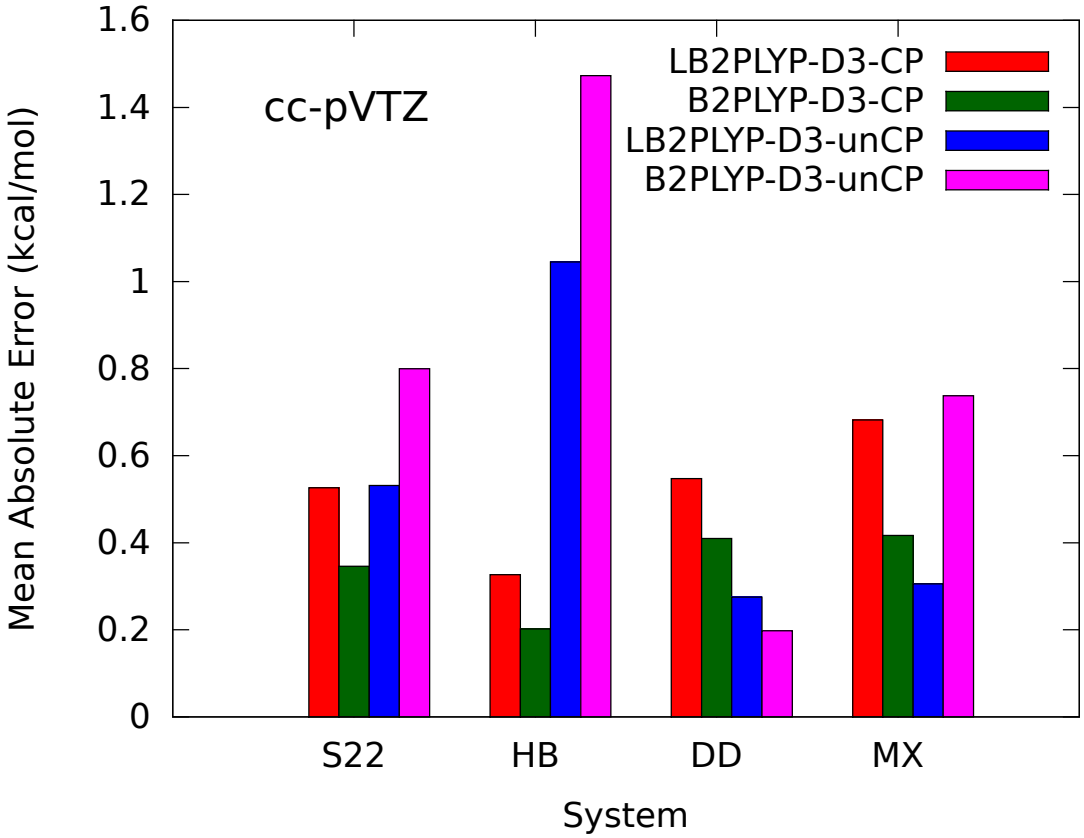


Figure 24: Mean absolute errors vs CCSD(T)/CBS reference values for S22 and the hydrogen bonding (HB), dispersion dominated (DD), and mixed interaction (MX) subsets of S22 computed in the cc-pVDZ (DZ) and cc-pVTZ (TZ) basis sets are presented in kcal mol^{-1}.

CHAPTER V

AUTOMATION OF WORKFLOWS INVOLVING F-ISAPT

5.1 Summary

Functional group and/or intramolecular symmetry-adapted perturbation theory (F-ISAPT) require tedious input files to begin the computation and perform the subsequent analysis. The following steps are required to form these inputs: A) monomers must be separated from the supersystem representation of a Protein Data Bank file or an XYZ file into the appropriate input sections and B) the fragments of each monomer must be specified by listing the fragment name and associated atom indices. This is easy to do for systems that are less than 30 atoms but quickly becomes intractable for systems that are any larger (i.e. 50+ atoms). I have developed Python-based tools for our collaborators at Bristol-Myers Squibb and students in our group that greatly facilitate F-ISAPT users with this process. One of the tools automatically generates all required fragment files and inputs given a Protein Data Bank file. Also, a visual representation of the generated fragmentation scheme is made for viewing with the PyMOL molecular visualization software. Finally, the other tool uses PyMOL as an interface for selecting custom fragmentation schemes and generates all required inputs.

5.2 Motivation

For all F-ISAPT computations, there are two rather tedious steps to setting up the input associated input files:

1. Separate the geometry into monomers for the main input file
2. Separate the monomers into functional groups for 'fA.dat' and 'fB.dat' files

The first item is a standard requirement for SAPT computations; however, this can be complicated if starting with a geometry that has monomers with "overlapping" atoms. Here, "overlapping" refers to the atoms of a geometry file not being ordered with respect

of the monomer which encompasses each atom. For the above scenario, a standard search algorithm, such as breadth-first search, can be employed to group the atoms by monomer. However, when explicit solvent is required for an ISAPT computation, there will be many monomers that will need to be extracted. Having each monomer separated may be of little use when there are many monomers. Additionally, in transition state geometries, there may not be a clear separation of monomers. It would be of value to have a tool that is able to specify the desired monomers for an input through a simple, intuitive mechanism.

The second item of separating monomers into functional-groups is specific to F-SAPT analysis. The input files for this type of analysis are structured like the following example for methanol:

```
OH 1 3
Methyl 2 4 5 6
```

Here, 'OH' represents a hydroxyl fragment of methanol, and 'Methyl' represents the CH3 fragment of the methanol monomer. This example is easy enough to do by hand but captures the most tedious part of creating this type of input. The atoms for the fragments do not necessarily retain any particular order in the input geometry. As a result, the atoms associated with each functional group have discontiguous atom ranks. Notice that the 'OH' atom numbers skip '2', and the 'Methyl' atom numbers skip '3'. This is easy enough to spot for a small input geometry in a molecular viewer with the atom ranks shown; however, this quickly becomes tedious in practice to designate even a few functional groups with input geometries of more than 50 atoms are used.

5.3 Solutions

Two solutions have been drafted to solve this problem. One solution pertains to the special case that the user has a geometry from the Protein Data Bank *and* wishes to fragment the system in such a way that is consistent with that of Parrish *et al., Chem. Eur. J* 2017, **23**, 7887. The other solution pertains to the more general case that a user has an XYZ file of Cartesian coordinates that are not in any particular order with respect to the atoms.

5.3.1 Protein Data Bank

As described in the study on the nature of the binding differences of chloro vs methyl aryl substituted fact Xa inhibitors,[52] the protein model system may be fragmented into peptide bonds, N-terminus caps, C-terminus caps, and amino acid sidechains. Additionally, free amino acids should be single fragments, and the ligand may be a single fragment as well.

The information needed to combine all fragments in this desired way is given in a properly formatted PDB file. Atoms belonging to peptide bonds have an atom type of 'N', 'C', 'O', and 'H'. The peptide bonds the identified atoms belong to can be realized by performing a breadth-first search (BFS) procedure to gather the peptide bond fragments. Amino acid residues, C-termini, and N-termini all have unique residue numbers and can be grouped into fragments on the basis of this ID. Finally, atoms not belonging to the protein are labeled as heteroatoms, 'HETATM'. These atoms belong to solvent and ligand molecules. By default, solvent molecules are included in fragment A with the protein, and ligand molecules as a whole are considered fragment B and are not fragmented more finely.

5.3.1.1 Instructions

Input: well-formed PDB, option for which monomer the solvent belongs to (default: solv_mon = "A")

Output: Psi4 input file, fA.dat and fB.dat fragmentation files

To use (in PyMOL):

1. Open PDB in PyMOL
2. Run aa_chopper.py in PyMOL (Note: this step can also be achieved by opening the script and .pdb simultaneously, perhaps through aliasing)
3. Type 'chop', or 'chop(solv_mon="C")' to put solvent in monomer C

This will produce:

1. A colored fragmentation scheme along with fragment names
2. A working input file with some sensible defaults

3. fsapt/fA.dat and fsapt/fB.dat files for post-analysis

    - Warning: it will overwrite any fA.dat or fB.dat files living in fsapt/ in the current working directory
    - PDB file must have '.pdb' extension
    - Input file produced will have the same name as PDB, with extension '.in'
    - Assumes there is only one ligand, and it is its own fragment
    - Assumes solvent does not carry charge and belongs to monomer A
    - Assumes well-formed PDB supplied with correct atom types for atoms participating in peptide bond

4. Fragment names will be as follows:

    - N-terminus caps: Chain_AminoAcidAbbreviation+ResidueNumber_NTC
    - C-terminus caps: Chain_AminoAcidAbbreviation+ResidueNumber_CTC
    - Sidechain caps: Chain_AminoAcidAbbreviation+ResidueNumber_CS
    - Sidechains: Chain_AminoAcidAbbreviation+ResidueNumber_SC
    - Peptide bonds: Chain_ResidueNumberOfAA1_ResidueNumberOfAA2_PEPT
    - Free Amino Acids: Chain_AminoAcidAbbreviation+ResidueNumber_FREE
    - Ligand: LIG_ResidueName_ResidueNumber
    - Solvent: SOLV_ResidueName_ResidueNumber

5. A PyMOL file named the same as the given PDB file, but with .pml extension

    - This will produce the same visualization as described above when opened with PyMOL
    - This can be done by typing 'pymol -l NAME_OF_FILE.pymol'

To use (in terminal):

1. Type 'python aa_chopper.py FILENAME.pdb', or 'python aa_chopper.py solv_mon = "C" FILENAME.pdb' to put solvent in monomer C

This will produce all of the same files described above.

5.3.2 XYZ

In the case that only the XYZ coordinates and atoms are provided, custom selections can be made through the PyMOL interface.[61] The process of fragmenting in this way occurs in the following steps:

- Select all border atoms of fragment
- Name fragment with '_A' or '_B' at the end of the fragment name
- Run 'fisapt' script

Border atoms in this context are defined as atoms the are in a user-desired fragment and connect to other parts of the molecule that are not included in this fragment.

This will produce the fragment files 'fA.dat' and 'fB.dat' and the input file. The software that takes care of this is broken into a part that grabs the input from the PyMOL interface and a part that produces the desired files given the user input.

The algorithm that produces the desired input files is represented by the psuedocode below.

Algorithm 1 Producing the desired fragments from a user's input of border atoms.

Require: Remove all selected atoms from copy of geometry

for selection ∈ selections **do**

Add atoms of selection back to copy of geometry

BFS to find fragments

Find fragment containing atoms of selection

Add rest of fragment to selection and save selection

Remove atoms of selection for copy of geometry

end for

return selections

5.3.3 PyMOL Interface

The PyMOL interface can be used to gather user input *via* its graphical user interface, which eliminates the tedious aspect of select atoms by the atom ID. The details of PyMOL's application programming interface can be difficult to find on the internet, but below are a few collected examples of commands that were instrumental in building these tools.

- 'from pymol import cmd' - grabs the command 'cmd' object that can be used to perform scriptable PyMOL operations
- 'cmd.select("("+frag+")", selection)' - adds the atom ids in 'selection' to the selection named 'frag'
- 'cmd.set_color("color_"+frag, color)' - defines 'color_frag' by 'color', which is a three element list of floats between 0 and 1 representing RGB values
- 'cmd.color("grey", "("+frag+")") - colors fragment '(frag)' a PyMOL defined color "grey"
- 'cmd.color("color_"+frag, "("+frag+")")' - colors fragment '(frag)' the color defined by 'color_frag'
- 'fil_name = cmd.get_names("all")[0]+".xyz"' - grabs file name, which is 0^{th} element of 'names'
- 'cmd.show("sticks", "all")' - shows stick representation of molecule
- 'frag_names = cmd.get_names("all")[1:]' - gets all of the selection names, skips 0^{th} element because that is geometry name
- 'cmd.iterate("("+name+")","stored.list.append((name,rank))")' - iterates over all atoms in geometry and returns name of atom and its rank on each iteration
- 'cmd.extend("fisapt",fisapt)' - names Python function 'fisapt', not in double quotes, "fisapt"

Below is an example instance of the tool being used with the PyMOL interface to designate F-ISAPT fragments in hexaphenylbenzene. The selection process just require the border atoms to be added as selections in the PyMOL graphical user interface, named 'Phenyl_A' and 'Phenyl_B' in this example. Additionally, the fragment C is determined by the border atoms in selection 'ISAPT_C', as shown in Figure 25. Once the tool is executed, the rest of the selection is filled in as shown in Figure 26. The fragmentation scheme shown is simultaneously written to the necessary input files to perform an F-SAPT analysis.

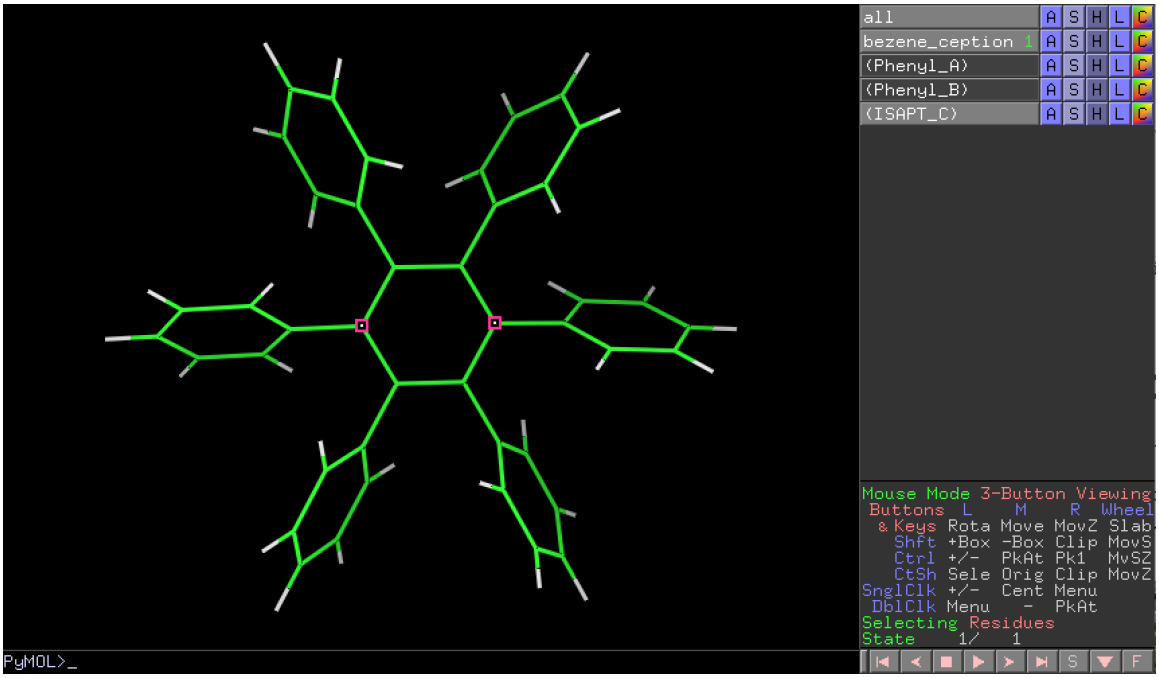


Figure 25: The necessary selections, shown by pink squares, to designate fragment C are shown. Only the border atoms, or atoms that connect to fragments not to be included in the selection, are necessary to include when using the input generation tool.

5.4 Conclusion

This code greatly simplifies the task of creating input files for the F-ISAPT analysis of large geometries and also reduces the chance for error. The problem of transcribing atom IDs from molecule viewers to input files has been reduced to selecting border atoms through PyMOL's interface. A visual representation of the fragmentation scheme chosen is provided each time a fragmentation is made and can be used to verify that the desired fragmentation

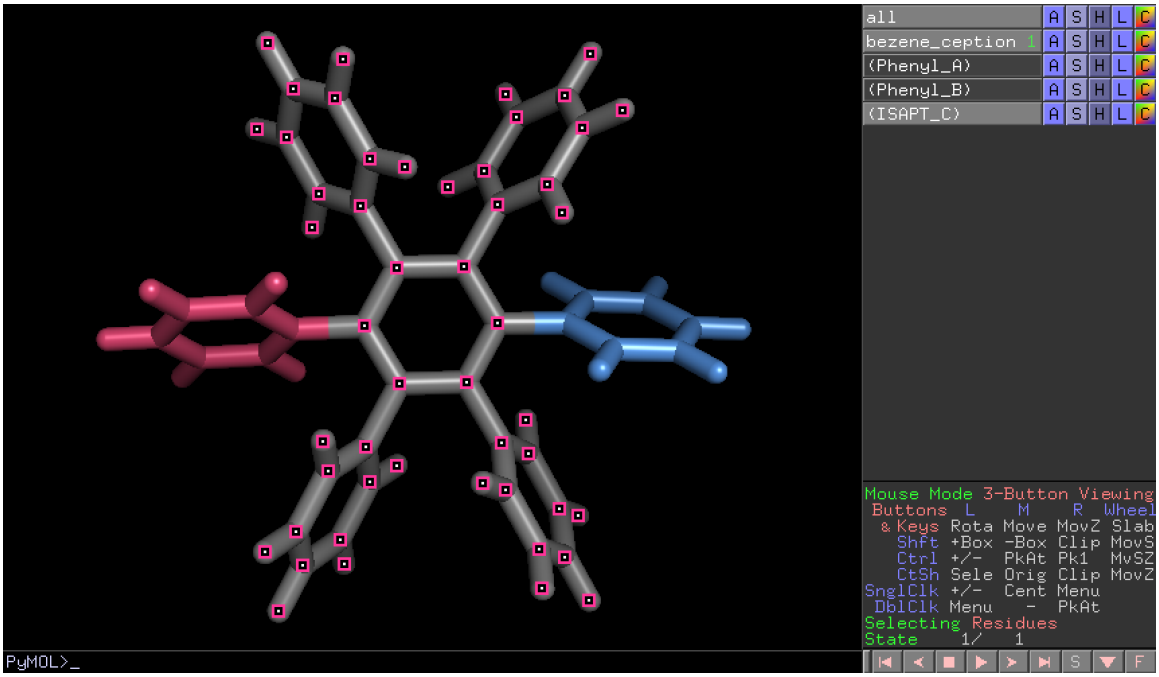


Figure 26: After the input generation tool is executed, the fragments are colored. The entire selection of fragment C is shown by the pink squares. Fragment A and B are colored by warm and cool colors, respectively.

scheme has been produced. The PyMOL commands necessary to build this tool are included in this text and make it possible to build more additional tools for input generation and data visualization in the future.

CHAPTER VI

CONCLUSIONS AND OUTLOOK

In this thesis, symmetry-adapted perturbation theory (SAPT) has been applied to two organocatalytic reactions: a proline-catalyzed intramolecular aldol reaction and an organoboron-catalyzed allyl addition to fluoroketones. The functional-group and atomic partitions of SAPT (F-SAPT and ASAPT) have been used to understand the non-covalent interactions involved in the transition states for these reactions in terms of electrostatics, exchange-repulsion, induction/polarization, and London dispersion. An understanding of these interactions and their effect on the relative product ratios of the reactions has potential to fuel rational catalyst design in the future. For instance, a pendant functional group could be added to a catalyst to make a racemic reaction selective. This work has sought to show that these partitions of SAPT can be used to rationalize experimental and other theoretical results in terms of directly quantified non-covalent interactions.

Our work on the proline-catalyzed intermolecular aldol reaction between benzaldehyde and cyclohexanone according to the Houk-List mechanism runs counter to previous attempts to rationalize the reaction outcome using chemical intuition alone. F-SAPT analysis shows that the $\mathrm{NCH}^{\delta+}\cdots^{\delta-}\mathrm{O{=}C}$ contact that was thought to be the primary source of stabilization for this reaction was in fact destabilizing and non-preferential to the transition state leading to the dominant product. This analysis was corroborated by modeling the Coulombic interaction of this contact in the transition state with natural population analysis (NPA) charges. The NPA charges showed that the nitrogen of pyrrolidinyl and oxygen of benzaldehyde, which are in close contact in the transition state, both carry large negative charges and make the contact destabilizing overall. Additional contacts, including a C–H/π interaction between the cyclohexene group of the enamine intermediate and the benzaldehyde phenyl ring and the hydrogen bonding interaction between an *ortho*-hydrogen of the phenyl and an oxygen of the carboxylic acid, were probed. F-SAPT confirmed the presence of both

of these interactions; however, the C–H/π interaction between the cyclohexene and phenyl groups was shown to be destabilizing in all cases. Only the hydrogen bonding interaction between the *ortho*-hydrogen of the phenyl and oxygen of the carboxylic acid was confirmed as preferentially stabilizing for the transition states leading to the major products in the Houk-List mechanism.

F-SAPT and ASAPT were used to assess contacts in organoboron-catalyzed allyl additions to fluoroketones that were hypothesized to contribute to the enantioselectivity of the reaction. ASAPT is in agreement for all atom-atom contacts as stabilizing or destabilizing as suggested by the authors of the study. The primary contact of interest was a hydrogen bonding interaction between a fluorine of the fluoroketone and an ammonium proton of the organoboron catalyst. ASAPT shows that this contact was indeed present as evidenced by its pronounced electrostatic stabilization, and was preferentially stabilizing the transition states postulated to be stabilized by this contact in the literature. F-SAPT analysis was also used to reveal a π-π interaction that was thought to be destabilizing. Due to the proximity of the phenyl groups and their reduced conjugation in the transition state, this contact was previously thought to present a source of steric hindrance. However, F-SAPT showed substantially stabilizing electrostatic and dispersion interactions which contribute to an overall stabilizing interaction.

These applications of the F-SAPT and ASAPT partitions to organocatalysis have pioneered the way for future studies of this nature. Exchange-scaling has been added to these methods and shown to track well with CCSD(T)/CBS relative interaction energies, within 1–4 kcal mol^{-1} out of total interaction energies of 37–69 kcal mol^{-1}. Our results for the fluoroketones fit chemical intuition. For the Houk-List mechanism, our results contradicted previous chemical intuition, but in retrospect and after additional atomic charge analysis, they also seem very chemically reasonable. Hence, these proof-of-concept studies have shown that we have a method that is both chemically reasonable and that quantitatively reproduces high-level quantum mechanical interaction energies to a reasonable accuracy. Future studies can confidently apply the methodology tested here to reactions which have not been probed experimentally.

Software tools have also been produced to simplify the process of input generation for both functional-group and intramolecular SAPT (F-ISAPT) methods, which aided in the studies mentioned above. These tools include software that can automatically produce all inputs required from a protein data bank file and software that allows the user to make custom selections through the PyMOL interface. In the future, these tools could be combined into more a general software that can serve any molecular file type. This could be achieved by adding a feature to automatically recognize functional groups, such as the algorithm for identifying functional groups of organic molecules described by Ertl.[19] Such a program could automatically generate fragmentation schemes without requiring any user input. This is desirable for performing future organocatalysis research and at drug companies that are looking to use F-ISAPT for high throughput work flows. Additionally, the F-SAPT results that the fragmentation schemes yield could be plotted to automate the path from input to analysis.

The local B2PLYP studies of this thesis aimed to probe the viability of substituting local MP2 for canonical MP2 in double hybrid density functional theory. Such a substitution would make B2PLYP applicable to larger systems by reducing the computational cost of canonical MP2. The study presented here tests local B2PLYP's ability to reproduce interaction energies for the S22 database. Local B2PLYP nearly reproduces B2PLYP for the entire test set by predicting average interaction differences of 0.65 kcal mol^{-1}. The next step in completing this assessment would require testing larger systems and systems that may not be easy to localize. Large protein-drug complexes or extended aromatic systems would certainly challenge local B2PLYP's ability to reproduce B2PLYP.

It is important to note that the applications of F-SAPT and ASAPT to organocatalysis presented here, which are the primary focus of this thesis, entirely rely on analysis of non-covalent interactions in the transition state. Non-covalent interactions can stabilize transition states and effectively lower a reaction barrier height relative to the reactants in infinite separation. However, stabilizing interactions between reactants or a reactant and catalyst could lead to a pre-reactive complex that would effectively raise a barrier height by lowering the energy of the starting reactants. For our analyses, we have made

the assumption that the lowest energy configuration of the reactants is the reactants in isolation. However, if some transition states being compared were reached through different pre-reactive complexes (or no pre-reactive complex at all), the computed relative barrier heights would not be able to be fairly compared. For future studies, it would be interesting to analyze the intrinsic reaction coordinates leading to each transition state of selective reactions, and obtain a SAPT analysis along the reaction path. This would ensure that relative barrier heights are fairly compared and would also reveal how the non-covalent interactions change during the course of the reaction. Additionally, the studies presented here exclusively used a gas phase model. Accounting for solvent in our models would likely screen, or reduce, the magnitude of the non-covalent interactions. Currently, there is not an implicit solvent model for SAPT, which could be developed in the future. This means that explicit solvent molecules would have to be added to the transition state geometry in order to incorporate any effects of solvent. In addition to screening, explicit solvent molecules in the transition state structures could result in unexpected non-covalent interactions that were entirely neglected in our studies.

Another potential point of interest would be exploring the effect of multipoles introduced by the F-SAPT fragmentation procedure. The simplest example of this effect is the fluorine dimer in a parallel geometry. When the fluorine atoms of each diatomic molecule are taken as separate fragments, F-SAPT analysis with the jun-cc-pVDZ basis shows that the fluorines diagonal to each other have stabilizing electrostatic interactions (-1.67 kcal mol^{-1} at 4.0 Å F_2 center of mass separation), and the fluorines directly across from each other have destabilizing electrostatic interactions (1.74 kcal mol^{-1} at 4.0 Å F_2 center of mass separation). This is presumably due to the analysis being performed on fragments that have electron densities representative of their molecular environment, but it is unclear how much of these interactions may be an artifact of the F-SAPT partitioning scheme.

Finally, there is a wealth of other reactions where non-covalent interactions are suspected to determine the selectivity.[32,42,44,57,75] It is the hope of the author that the tools and methodologies of this thesis will be used to provide SAPT based analyses of these and other organocatalytic reactions.

APPENDIX A

EXAMINATION OF NON-COVALENT INTERACTIONS IN THE HOUK-LIST MECHANISM

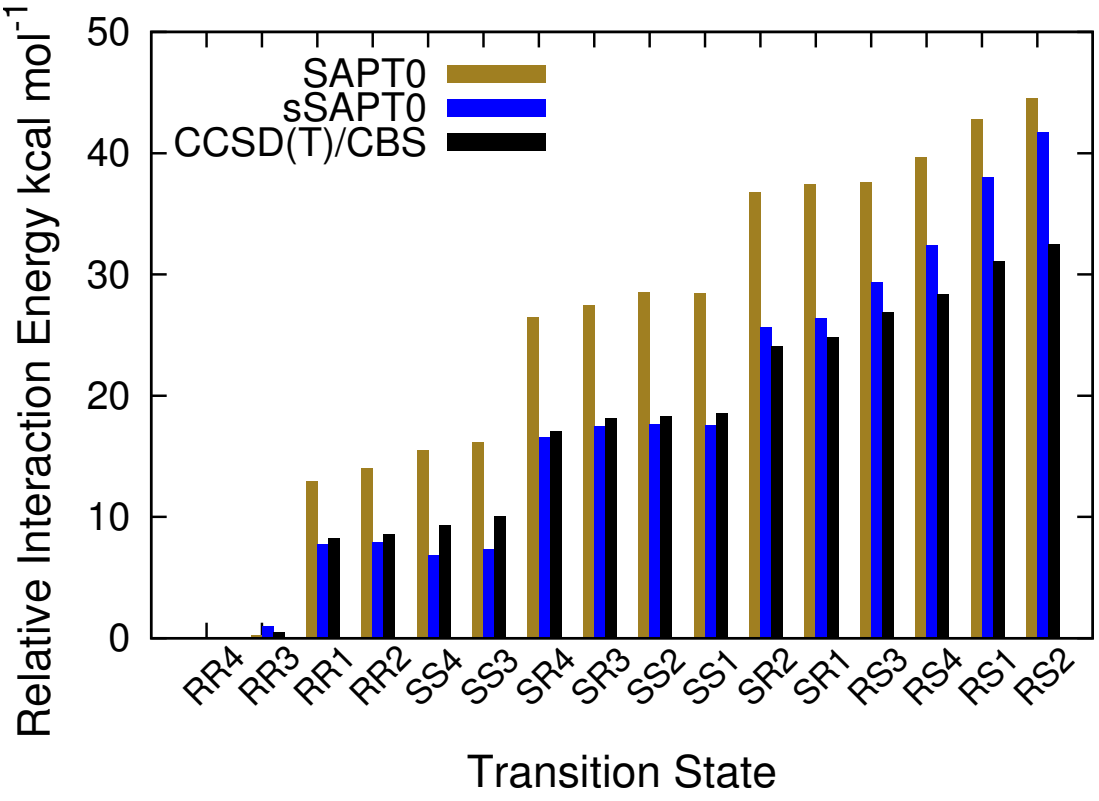


Figure 27: Estimated CCSD(T)/CBS, SAPT0/jun-cc-pVDZ, sSAPT0/jun-cc-pVDZ relative interaction energies, in increasing order of the CCSD(T)/CBS values. CCSD(T)/CBS values were estimated as DF-MP2/aug-cc-pV(T,Q)Z + $\delta_{\mathrm{MP2}}^{\mathrm{CCSD(T)}}$/cc-pVDZ.

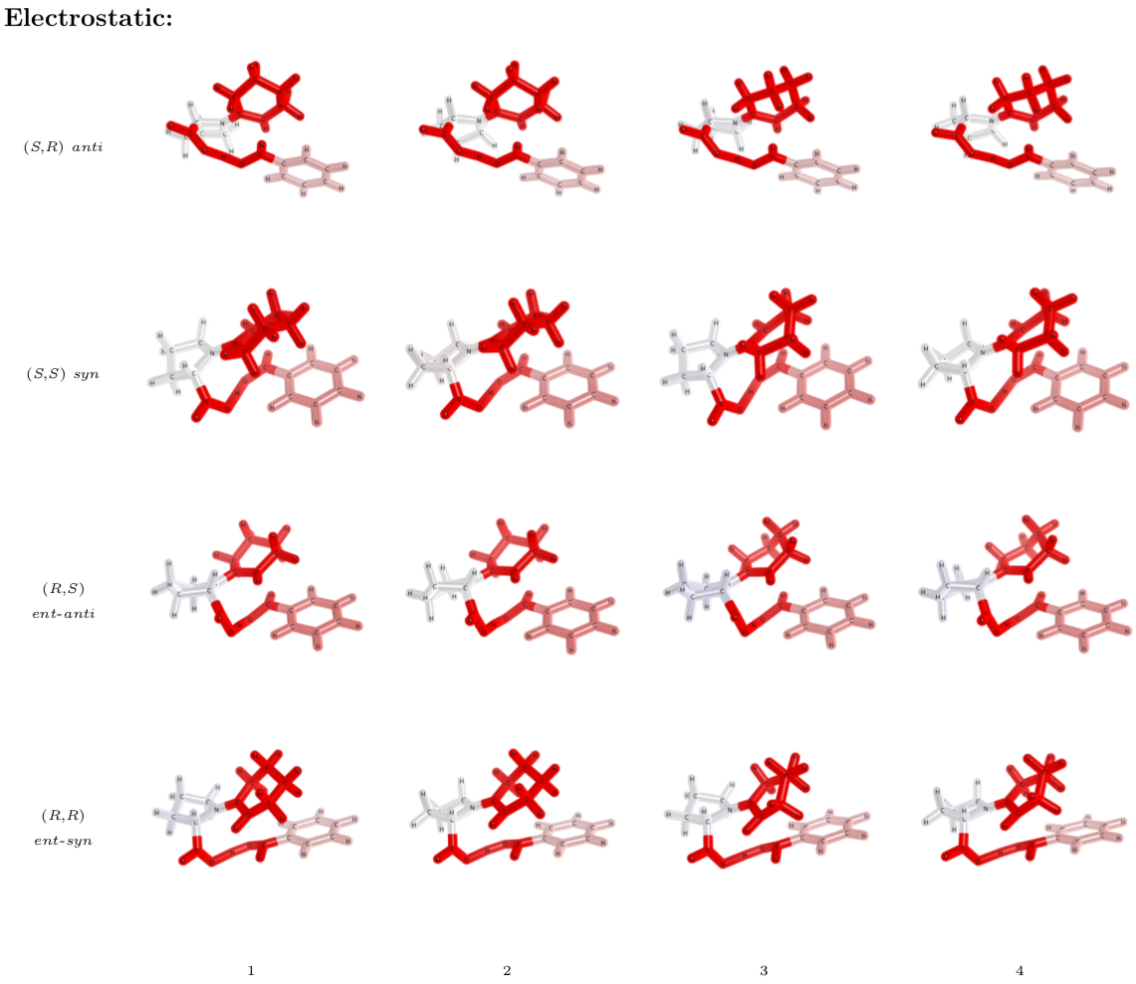


Figure 28: The total electrostatic interaction, as predicted by F-SAPT0/jun-cc-pVDZ, between each fragment and the other monomer is displayed for all considered transition states. Each row consists of transition states leading to the (*S*,*R*) *anti*, (*S*,*S*) *syn*, (*R*,*S*) *ent-anti*, and (*R*,*R*) *ent-syn* transition states respectively. The columns consist of the possible conformations (1,2,3, and 4 respectively) of each transition state. The color scale for the images becomes saturated at $\pm$ 25 kcal mol^{-1}.

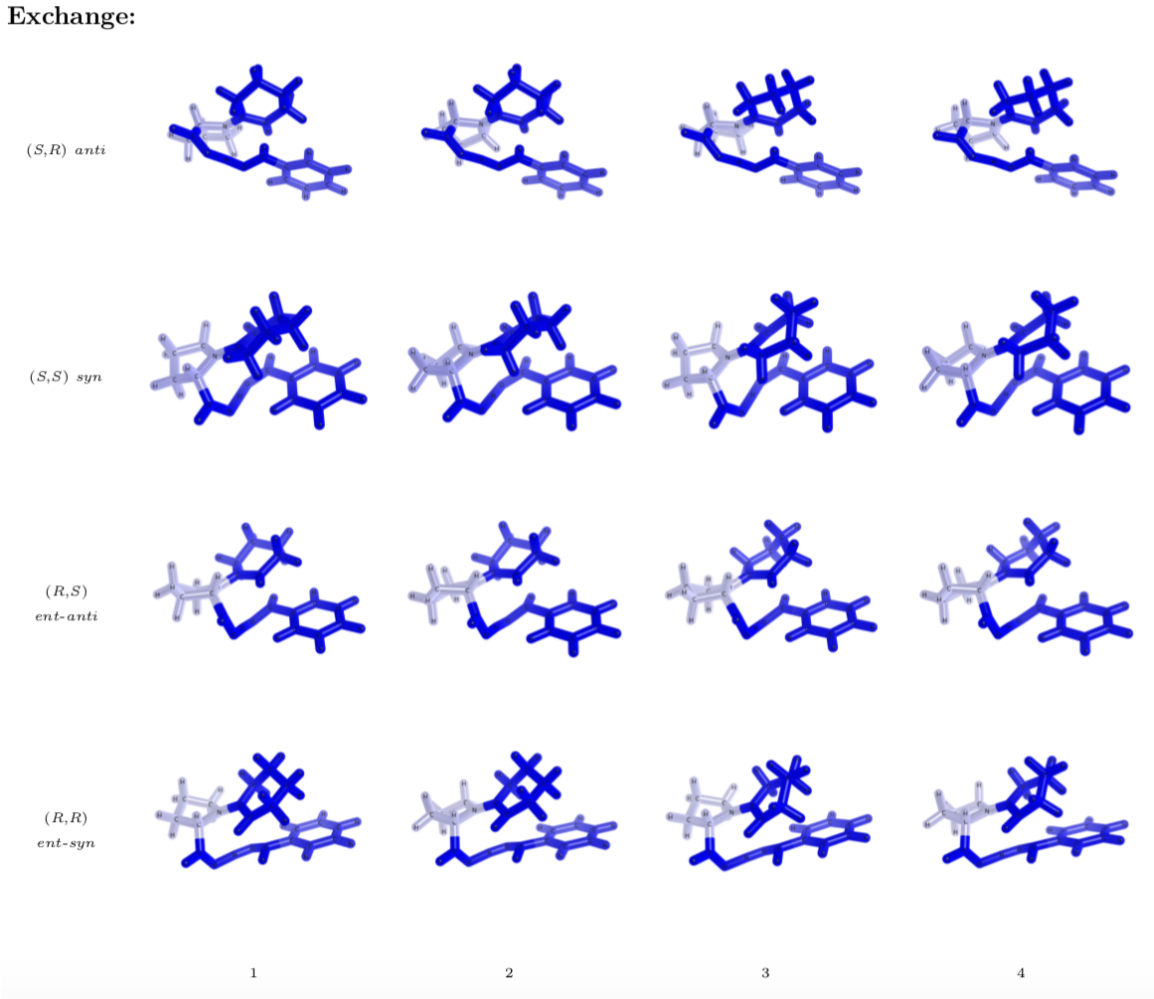


Figure 29: The total exchange interaction, as predicted by F-SAPT0/jun-cc-pVDZ, of each fragment with the other monomer is displayed for all considered transition states. Each row consists of transition states leading to the (*S*,*R*) *anti*, (*S*,*S*) *syn*, (*R*,*S*) *ent-anti*, and (*R*,*R*) *ent-syn* transition states respectively. The columns consist of the possible conformations (1,2,3, and 4 respectively) of each transition state. The color scale for the images becomes saturated at ± 25 kcal mol^{-1}.

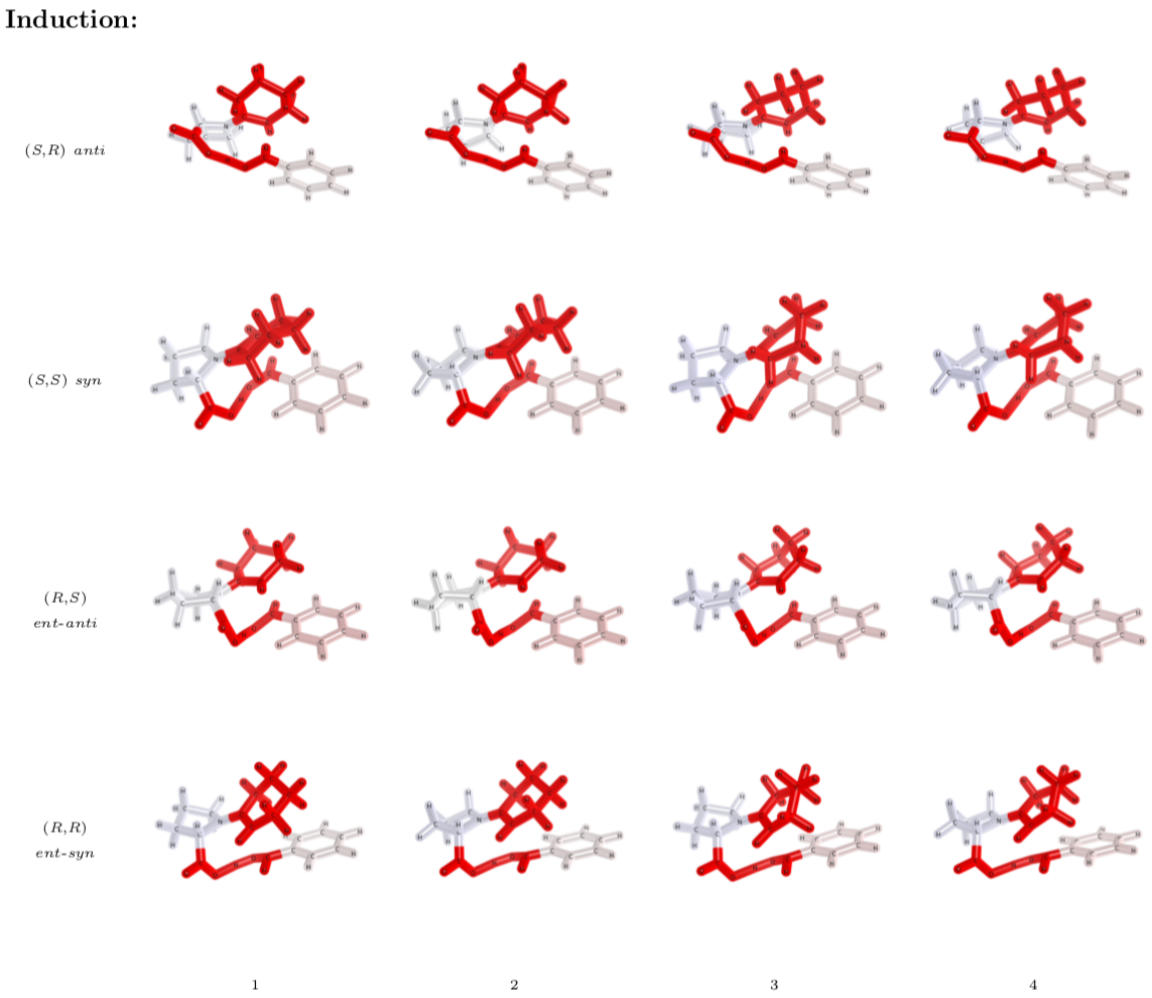


Figure 30: The total induction term, as predicted by F-SAPT0/jun-cc-pVDZ, for each fragment interacting with the other monomer is displayed for all considered transition states. Each row consists of transition states leading to the (*S*,*R*) *anti*, (*S*,*S*) *syn*, (*R*,*S*) *ent-anti*, and (*R*,*R*) *ent-syn* transition states respectively. The columns consist of the possible conformations (1,2,3, and 4 respectively) of each transition state. The color scale for the images becomes saturated at $\pm$ 25 kcal mol^{-1}.

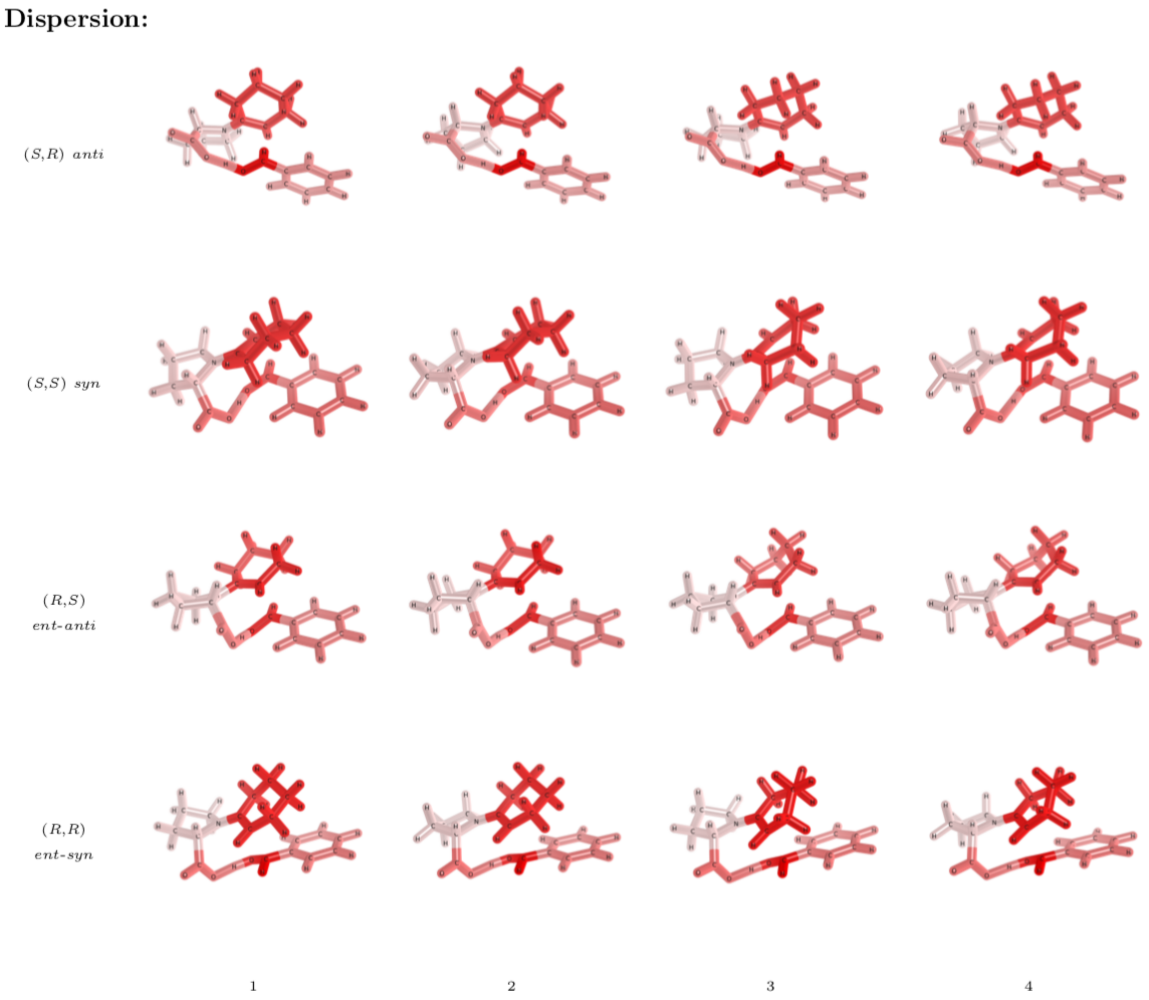


Figure 31: The total dispersion interaction, as predicted by F-SAPT0/jun-cc-pVDZ, between each functional group with the other monomer is displayed for all considered transition states. Each row consists of transition states leading to the (*S*,*R*) *anti*, (*S*,*S*) *syn*, (*R*,*S*) *ent-anti*, and (*R*,*R*) *ent-syn* transition states respectively. The columns consist of the possible conformations (1,2,3, and 4 respectively) of each transition state. The color scale for the images becomes saturated at ± 25 kcal mol^{-1}.

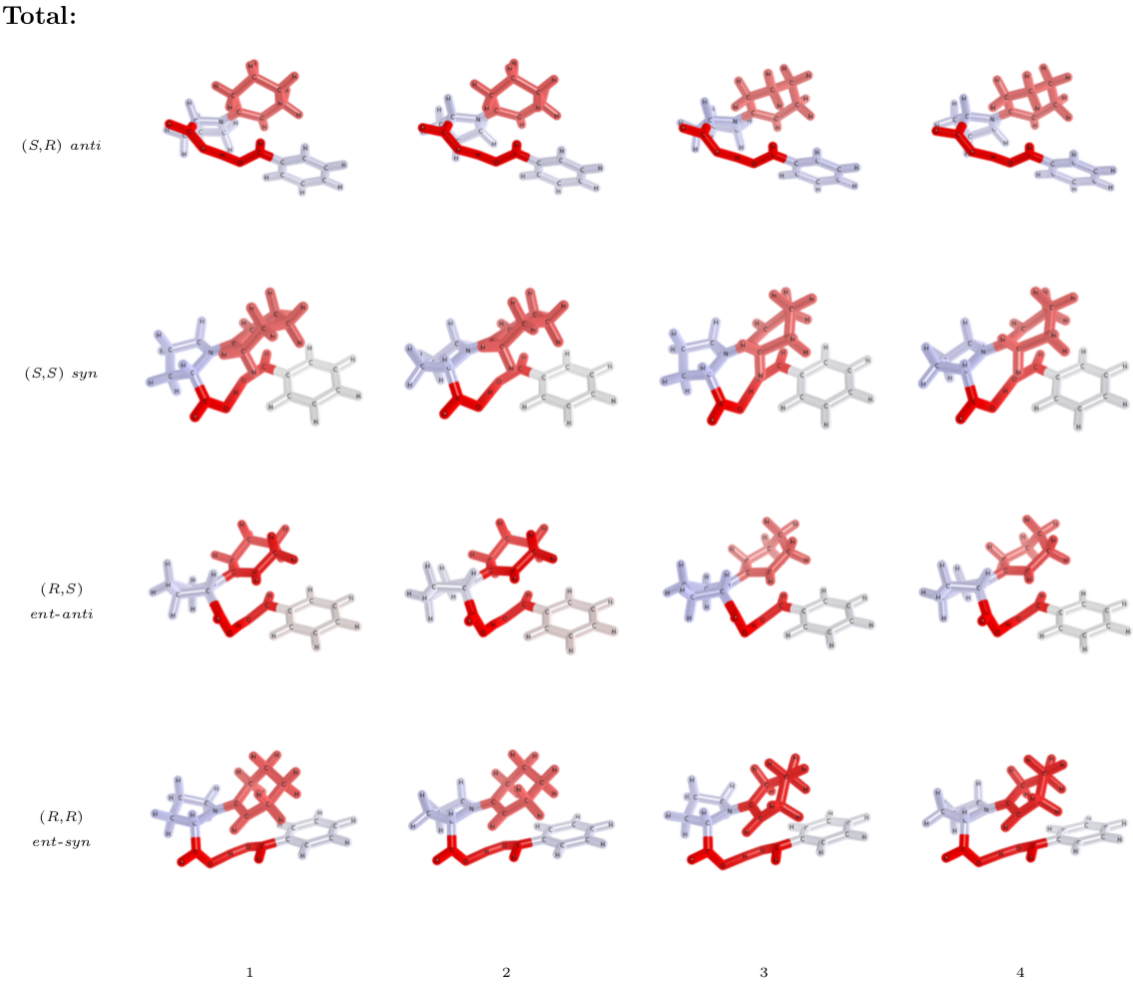


Figure 32: The total interaction, as predicted by F-SAPT0/jun-cc-pVDZ, of each functional group with the other monomer for all considered transition states. Each row consists of transition states leading to the (*S*,*R*) *anti*, (*S*,*S*) *syn*, (*R*,*S*) *ent-anti*, and (*R*,*R*) *ent-syn* transition states respectively. The columns consist of the possible conformations (1,2,3, and 4 respectively) of each transition state.

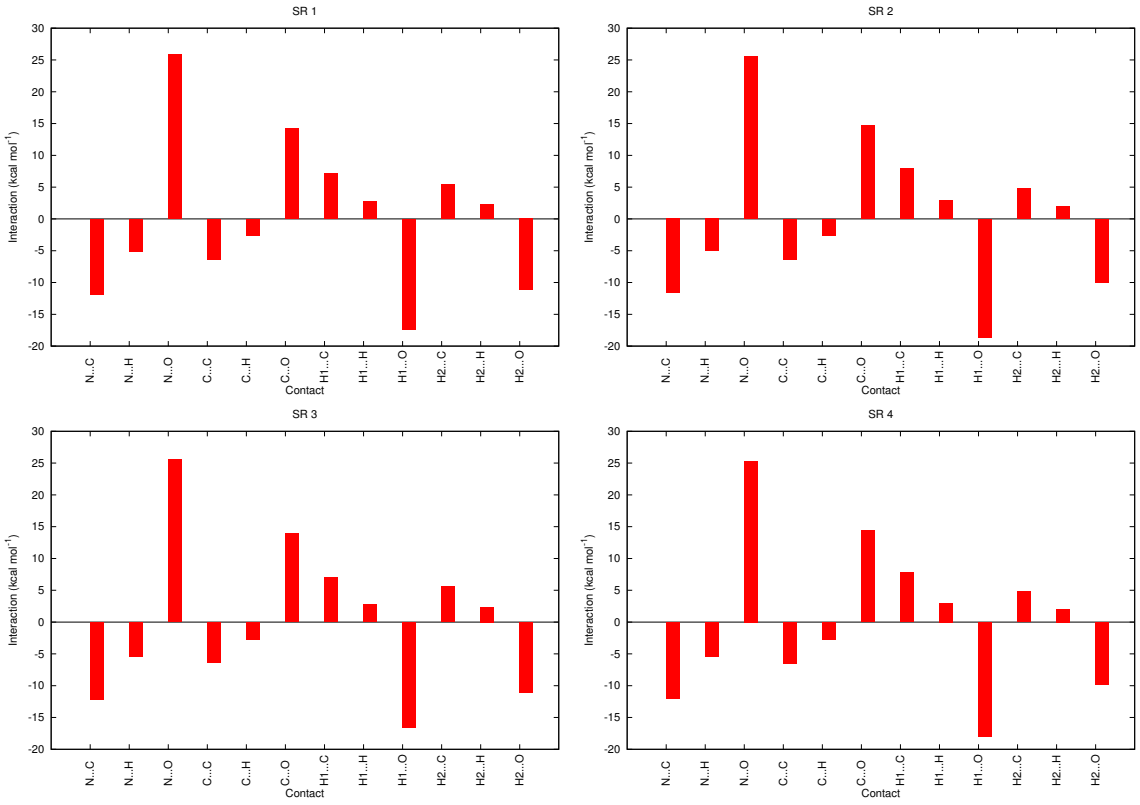


Figure 33: The classical Coulombic interaction between the partial charges on each of the atoms involved in the interaction between NCH_2 of the pyrrolidinyl fragment and the CHO of the aldehyde of fragment in kcal mol^{-1} for the (*S*,*R*) *anti* transition states. Here, H1 designates the proton of NCH_2 that is closer to the aldehyde, and H2 designates the proton that is farther from the aldehyde.

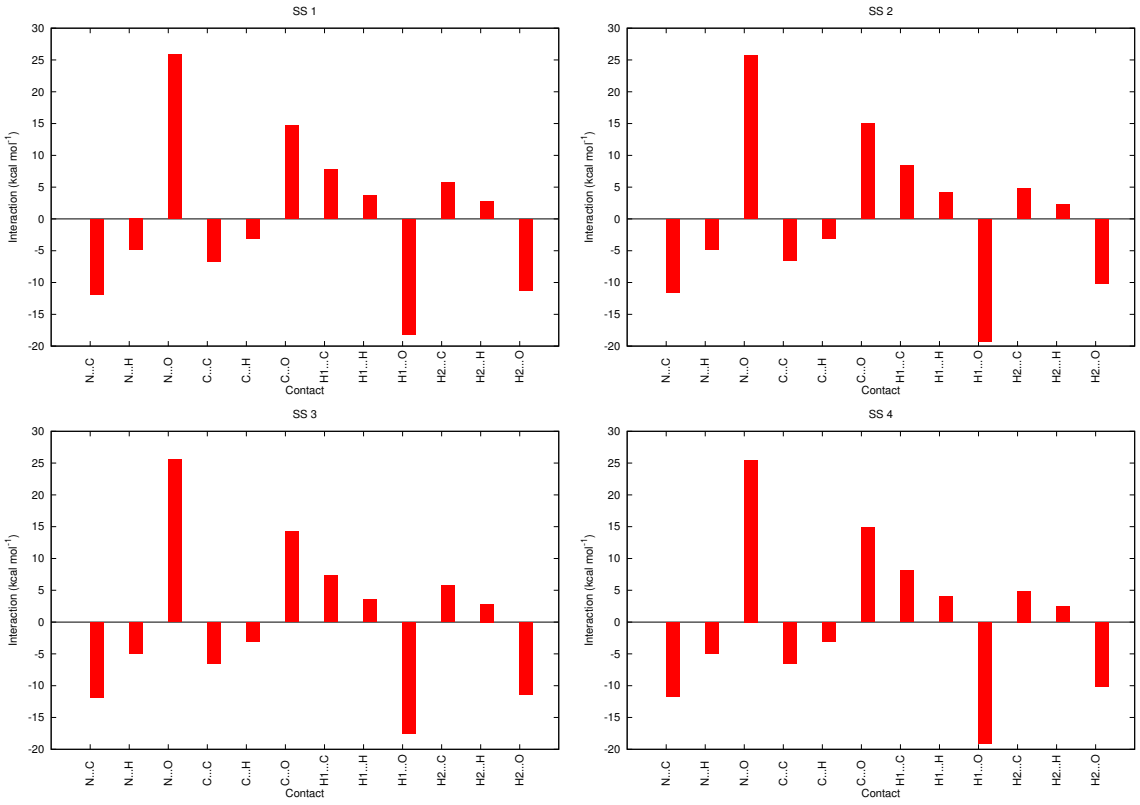


Figure 34: The classical Coulombic interaction between the partial charges on each of the atoms involved in the interaction between NCH_2 of the pyrrolidinyl fragment and the CHO of the aldehyde of fragment in kcal mol^{-1} for the (*S*,*S*) *syn* transition states. Here, H1 designates the proton of NCH_2 that is closer to the aldehyde, and H2 designates the proton that is farther from the aldehyde.

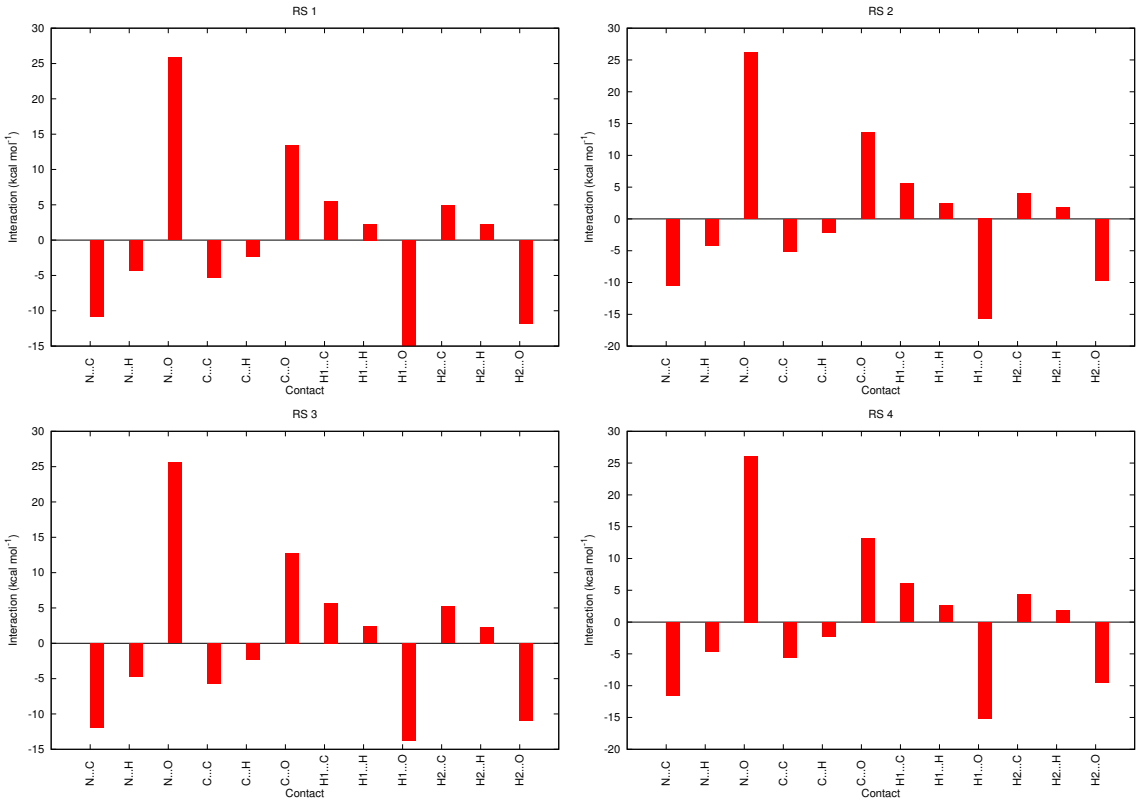


Figure 35: The classical Coulombic interaction between the partial charges on each of the atoms involved in the interaction between NCH_2 of the pyrrolidinyl fragment and the CHO of the aldehyde of fragment in kcal mol^{-1} for the (*R*,*S*) *ent-anti* transition states. Here, H1 designates the proton of NCH_2 that is closer to the aldehyde, and H2 designates the proton that is farther from the aldehyde.

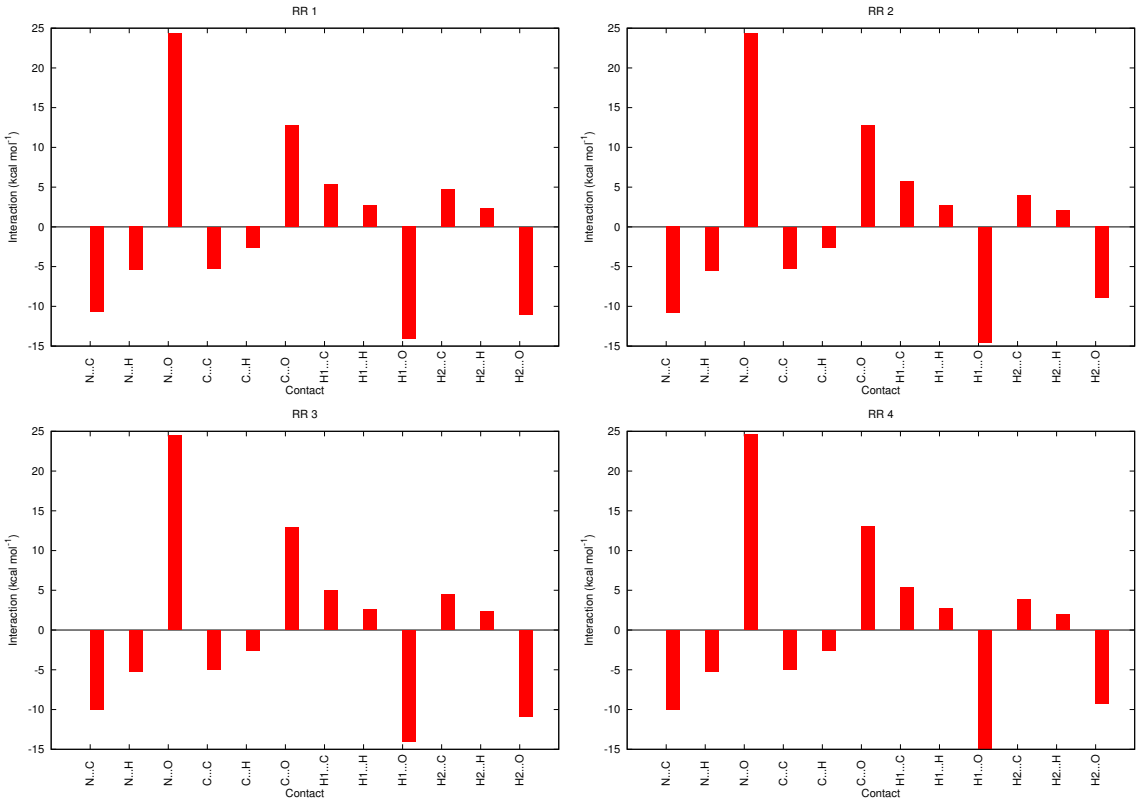


Figure 36: The classical Coulombic interaction between the partial charges on each of the atoms involved in the interaction between NCH_2 of the pyrrolidinyl fragment and the CHO of the aldehyde of fragment in kcal mol^{-1} for the (*R*,*R*) *ent-syn* transition states. Here, H1 designates the proton of NCH_2 that is closer to the aldehyde, and H2 designates the proton that is farther from the aldehyde.

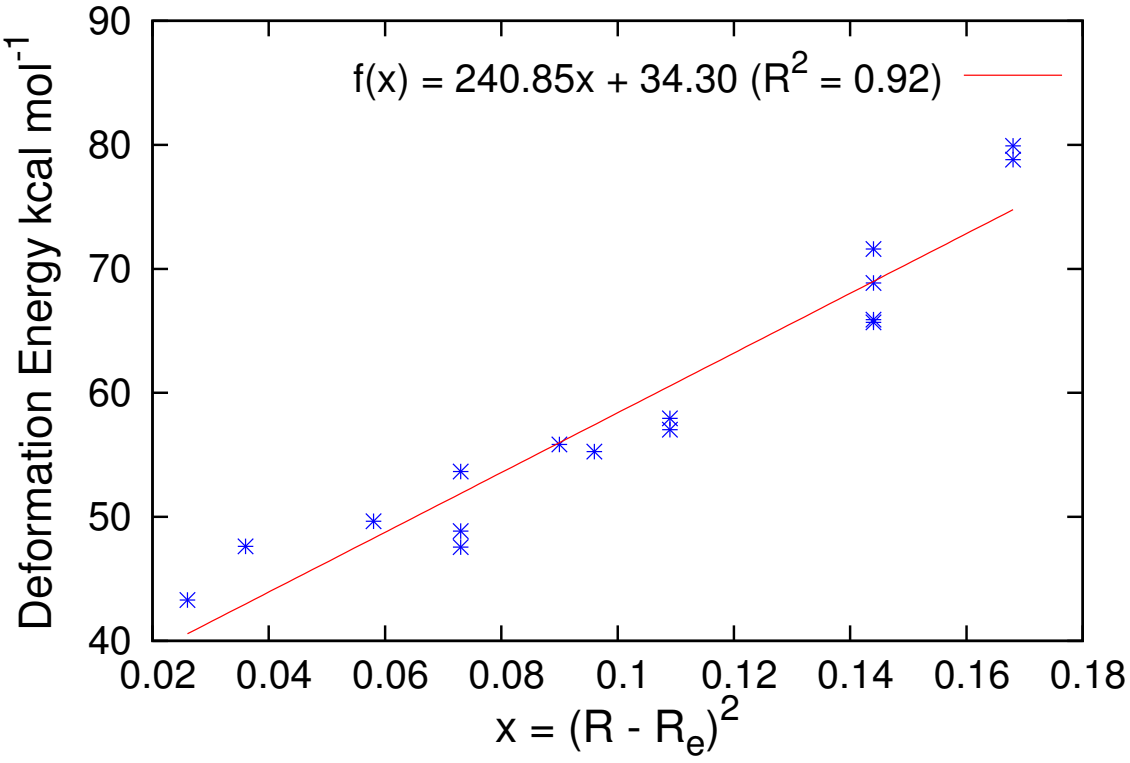


Figure 37: Deformation energies estimated at the CCSD(T)/CBS level are plotted against $(R - R_e)^2$, where R is the O–H bond length in the carboxylic acid group, and R_e is its value in the isolated enamine intermediate. CCSD(T)/CBS values were estimated as DF-MP2/aug-cc-pV(T,Q)Z + $\delta_{\mathrm{MP2}}^{\mathrm{CCSD(T)}}$/cc-pVDZ.

APPENDIX B

ANALYSIS OF TRANSITION STATE STABILIZATION BY NON-COVALENT INTERACTIONS IN THE ADDITION OF ORGANOBORON REAGENTS TO FLUOROKETONES

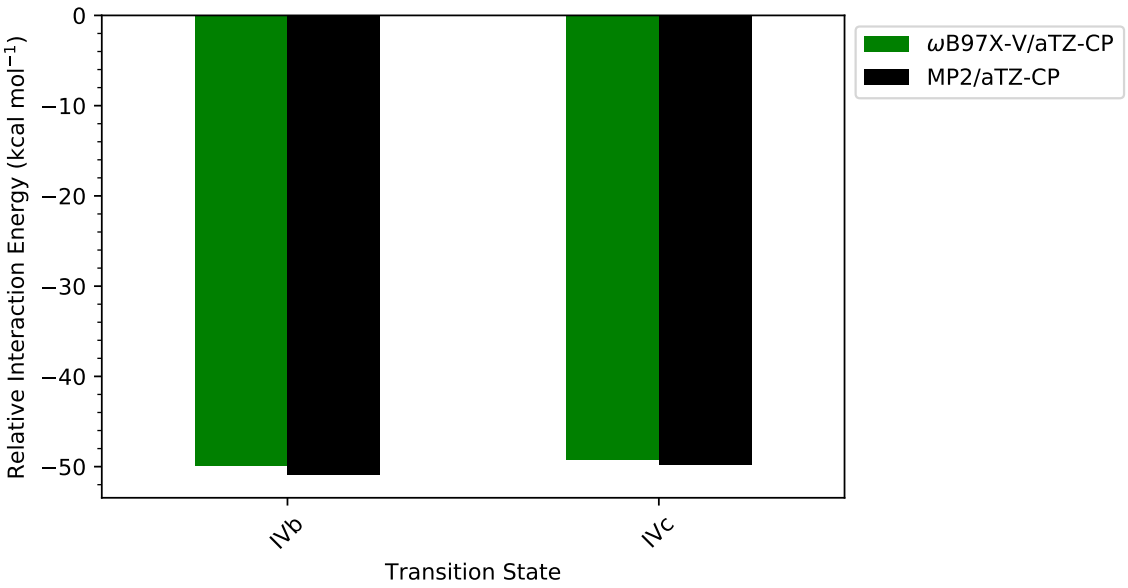


Figure 38: The chosen reference method, counterpoise-corrected MP2/aTZ was tested against counterpoise-corrected ωB97X-V/aTZ on two systems where π-π interactions were suspected, due to aligned π faces. MP2/aTZ estimates are within 1 kcal mol^{-1} of ωB97X-V/aTZ indicating no large overbinding.

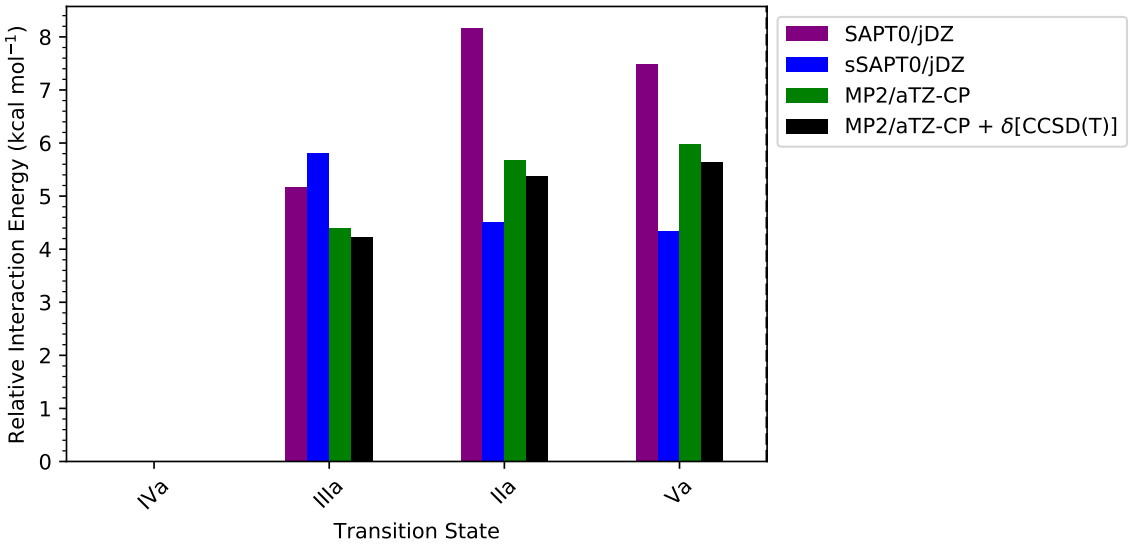


Figure 39: Interaction energies are plotted relative to the most negative interaction energy estimated within each model chemistry (SAPT0/jaDZ, exchange-scaled sSAPT0/jaDZ, counterpoise-corrected MP2/aTZ, counterpoise-corrected MP2/aTZ + counterpoise-corrected δ[CCSD(T)/DZ]).

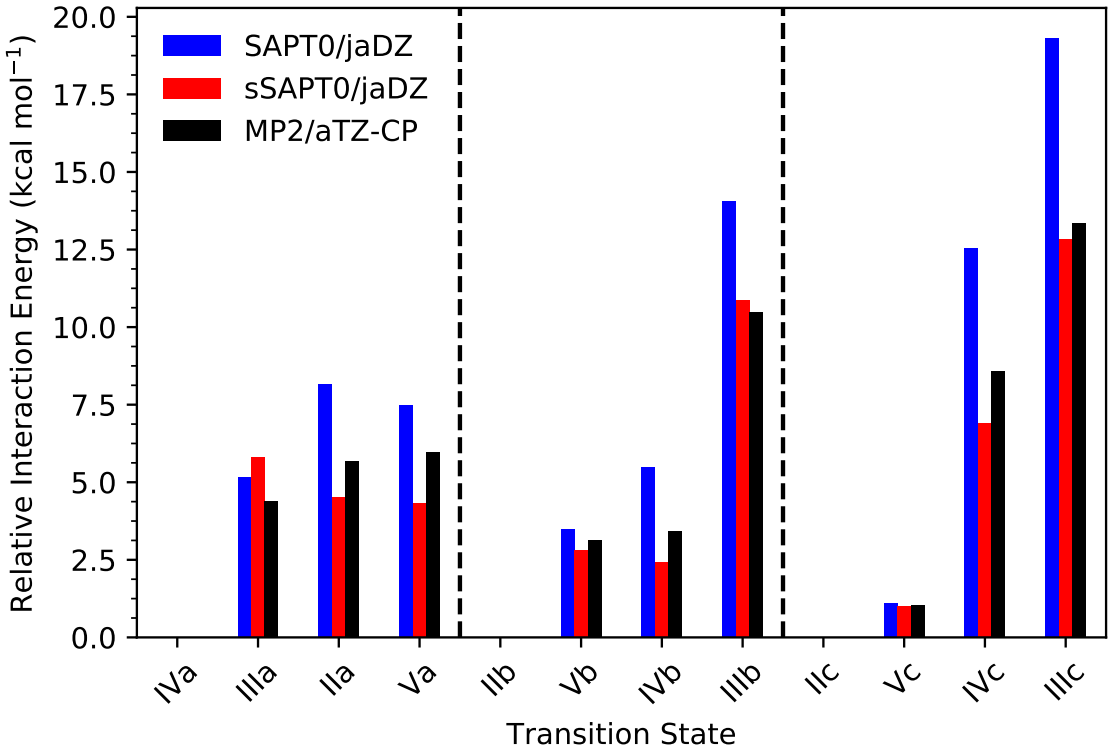


Figure 40: Interaction energies are plotted relative to the most negative interaction energy estimated within each model chemistry (SAPT0/jaDZ, exchange-scaled sSAPT0/jaDZ and MP2/aTZ).

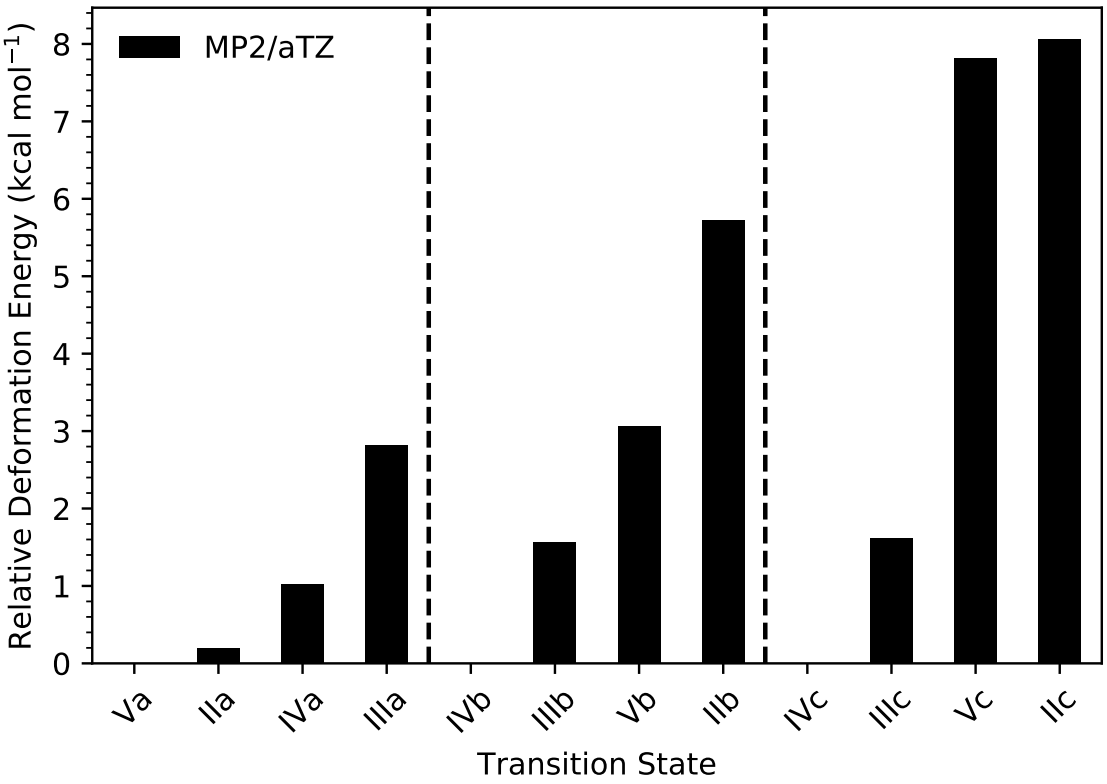


Figure 41: The relative deformation energy of each transition state, as computed with MP2/aTZ, is plotted with respect to the minimum deformation energy in each reaction family (A, B, and C).

REFERENCES

[1] Antony, J. and Grimme, S., "Is spin-component scaled second-order møller plesset perturbation theory an appropriate method for the study of noncovalent interactions in molecules?," *J. Phys. Chem. A*, vol. 111, pp. 4862–4868, 2007.

[2] Armstrong, A., Boto, R. A., Dingwall, P., Contreras-Garcia, J., Harvey, M. J., Mason, N. J., and Rzepa, H. S., "The houk-list transition states for organocatalytic mechanisms revisited," *Chem. Sci.*, vol. 5, pp. 2057–2071, 2014.

[3] Bahmanyar, S., Houk, K. N., Martin, H. J., and List, B., "Quantum mechanical predictions of the stereoselectivities of proline-catalyzed asymmetric intermolecular aldol reactions," *J. Am. Chem. Soc.*, vol. 125, pp. 2475–2479, 2003.

[4] Bakr, B. W. and Sherrill, C. D., "Analysis of transition state stabilization by non-covalent interactions in the Houk-List model of organocatalyzed intermolecular Aldol additions using functional-group symmetry-adapted perturbation theory," *Phys. Chem. Chem. Phys.*, vol. 18, pp. 10297–10308, 2016.

[5] Becke, A. D., "Density-functional exchange-energy approximations with correct asymptotic behavior," *Phys. Rev. A*, vol. 38, pp. 3098–3100, 1988.

[6] Becke, A. D., "Density-functional thermochemistry. iii. the role of exact exchange," *J. Chem. Phys.*, vol. 98, pp. 5648–5652, 1993.

[7] Becke, A. D., "A new mixing of Hartree-Fock and local density-functional theories," *J. Chem. Phys.*, vol. 98, pp. 1372–1377, 1993.

[8] Biczysko, M., Panek, P., Scalmani, G., Bloino, J., and Barone, V., "Harmonic and anharmonic vibrational frequency calculations with the double-hybrid b2plyp method: Analytic second derivatives and benchmark studies," *J. Chem. Theory Comput.*, vol. 6, no. 7, pp. 2115–2125, 2010.

[9] Boughton, J. W. and Pulay, P., "Comparison of the Boys and Pipek-Mezey localizations in the local correlation approach and automatic virtual basis selection," *J. Comput. Chem.*, vol. 14, no. 6, pp. 736–740, 1993.

[10] Boys, S. F., "Localized orbitals and localized adjustment functions," in *Quantum Theory of Atoms, Molecules, and the Solid State* (Löwdin, P.-O., ed.), pp. 253–262, New York: Academic Press, 1966.

[11] Boys, S. F. and Bernardi, F., "The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors," *Mol. Phys.*, vol. 19, no. 4, pp. 553–566, 1970.

[12] Burns, L. A., Vázquez-Mayagoitia, Á., Sumpter, B. G., and Sherrill, C. D., "Density-functional approaches to noncovalent interactions: A comparison of dispersion corrections (DFT-D), exchange-hole dipole moment (XDM) theory, and specialized functionals," *J. Chem. Phys.*, vol. 134, p. 084107, 2011.

[13] Contreras-García, J., Johnson, E. R., Keinan, S., Chaudret, R., Piquemal, J., Beratan, D. N., and Yang, W., "NCIPLOT: A program for plotting noncovalent interaction regions," *J. Chem. Theory Comput.*, vol. 7, pp. 625–632, 2011.

[14] DePrince, A. E. and Sherrill, C. D., "Accuracy and efficiency of coupled-cluster theory using density fitting/cholesky decomposition, frozen natural orbitals, and a t_1-transformed hamiltonian," *J. Chem. Theory Comput.*, vol. 9, pp. 2687–2696, 2013.

[15] Dunlap, B. I., Connolly, J. W. D., and Sabin, J. R., "Applicability of LCAO-X-alpha methods to molecules containing transition-metal atoms - nickel atom and nickel hydride," *Int. J. Quantum Chem. Symp.*, vol. 11, p. 81, 1977.

[16] Dunlap, B. I., Connolly, J. W. D., and Sabin, J. R., "On some approximations in applications of $X\alpha$ theory," *J. Chem. Phys.*, vol. 71, pp. 3396–3402, 1979.

[17] Dunning, T. H., "Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen," *J. Chem. Phys.*, vol. 90, pp. 1007–1023, 1989.

[18] East, A. L. L. and Allen, W. D., "The heat of formation of nco," *J. Chem. Phys.*, vol. 99, no. 6, pp. 4638–4650, 1993.

[19] Ertl, P., "An algorithm to identify functional groups in organic molecules," *Journal of Cheminformatics*, vol. 9, no. 1, p. 36, 2017.

[20] Ess, D. H. and Houk, K. N., "Distortion/interaction energy control of 1,3-dipolar cycloaddition reactivity," *J. Am. Chem. Soc.*, vol. 129, pp. 10646–10647, 2007.

[21] Fernández, I. and Bickelhaupt, F. M., "The activation strain model and molecular orbital theory: Understanding and designing chemical reactions," *Chem. Soc. Rev.*, vol. 43, pp. 4953–4967, 2014.

[22] Glendening, E. D., Badenhoop, J. K., Reed, A. E., Carpenter, J. E., Bohmann, J. A., Morales, C. M., and Weinhold, F., "NBO 5.0." Theoretical Chemistry Institute: University of Wisconsin, Madison, WI, 2001. http://www.chem.wisc.edu/∼nbo5.

[23] Grimme, S., "Semiempirical hybrid density functional with perturbative second-order correlation," *J. Chem. Phys.*, vol. 124, p. 034108, 2006.

[24] Grimme, S., Antony, J., Ehrlich, S., and Krieg, H., "A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu," *J. Chem. Phys.*, vol. 132, p. 154104, 2010.

[25] Halkier, A., Helgaker, T., Jørgensen, P., Klopper, W., Koch, H., Olsen, J., and Wilson, A. K., "Basis-set convergence in correlated calculations on Ne, N_2, and H_2O," *Chem. Phys. Lett.*, vol. 286, pp. 243–252, 1998.

[26] Hampel, C. and Werner, H.-J., "Local treatment of electron correlation in coupled cluster theory," *J. Chem. Phys.*, vol. 104, pp. 6286–6297, 1996.

[27] Hesselmann, A., Jansen, G., and Schütz, M., "Density-functional theory-symmetry-adapted intermolecular perturbation theory with density fitting: A new efficient method to study intermolecular interaction energies," *J. Chem. Phys.*, vol. 122, p. 014103, 2005.

[28] Hetzer, G., Pulay, P., and Werner, H.-J., "Multipole approximation of distant pair energies in local MP2 calculations," *Chem. Phys. Lett.*, vol. 290, pp. 143–149, 1998.

[29] Hetzer, G., *Multipolnäherungen für lokale Korrelationsverfahren.* PhD thesis, University of Stuttgart, Stuttgart, Germany, 2000.

[30] Hohenstein, E. G. and Sherrill, C. D., "Density fitting and Cholesky decomposition approximations in symmetry-adapted perturbation theory: Implementation and application to probe the nature of π-π interactions in linear acenes," *J. Chem. Phys.*, vol. 132, p. 184111, 2010.

[31] Hohenstein, E. G. and Sherrill, C. D., "Density fitting of intramonomer correlation effects in symmetry-adapted perturbation theory," *J. Chem. Phys.*, vol. 133, p. 014101, 2010.

[32] Holland, M. C., Berden, G., Oomens, J., Meijer, A. J. H. M., Schäfer, M., and Gilmour, R., "Infrared multiphoton dissociation spectroscopic analysis of noncovalent interactions in organocatalysis," *European Journal of Organic Chemistry*, vol. 2014, no. 26, pp. 5675–5680.

[33] Jeziorski, B., Moszynski, R., and Szalewicz, K., "Perturbation theory approach to intermolecular potential energy surfaces of van der Waals complexes," *Chem. Rev.*, vol. 94, pp. 1887–1930, 1994.

[34] Johnson, E. R., Keinan, S., Mori-Sanchez, P., Contreras-Garcia, J., Cohen, A. J., and Yang, W., "Revealing noncovalent interactions," *J. Am. Chem. Soc.*, vol. 132, pp. 6498–6506, 2010.

[35] Jurečka, P., Šponer, J., Černý, J., and Hobza, P., "Benchmark database of accurate (MP2 and CCSD(T) complete basis set limit) interaction energies of small model complexes, DNA base pairs, and amino acid pairs," *Phys. Chem. Chem. Phys.*, vol. 8, pp. 1985–1993, 2006.

[36] Kendall, R. A., Dunning, T. H., and Harrison, R. J., "Electron affinities of the first-row atoms revisited. systematic basis sets and wave functions," *J. Chem. Phys.*, vol. 96, pp. 6796–6806, 1992.

[37] Kohn, W. and Sherrill, C. D., "Editorial: Reflections on fifty years of density functional theory," *J. Chem. Phys.*, vol. 140, p. 18A201, 2014.

[38] Krenske, E. H. and Houk, K. N., "Aromatic interactions as control elements in stereoselective organic reactions," *Acc. Chem. Res.*, vol. 46, pp. 979–989, 2013.

[39] Lao, K. U. and Herbert, J. M., "Breakdown of the single-exchange approximation in third-order symmetry-adapted perturbation theory," *J. Phys. Chem. A*, vol. 116, pp. 3042–3047, 2012.

[40] Lee, K. A., Silverio, D., Torker, S., Robbins, D. W., Haeffner, F., van der Mei, F. W., and Hoveyda, A. H., "Catalytic enantioselective addition of organoboron reagents to fluoroketones controlled by electrostatic interactions," vol. 8, pp. 768–777, 2016.

[41] Lee, T. J., Remington, R. B., Yamaguchi, Y., and Schaefer, H. F., "The effects of triple and quadruple excitations in configuration interaction procedures for the quantum mechanical prediction of molecular properties," *J. Chem. Phys.*, vol. 89, pp. 408–422, 1988.

[42] Lin, S. and Jacobsen, E., "Thiourea-catalysed ring opening of episulfonium ions with indole derivatives by means of stabilizing non-covalent interactions," vol. 4, pp. 817–824, 10 2012.

[43] Maji, R. and Wheeler, S. E., "Importance of electrostatic effects in the stereoselectivity of nhc-catalyzed kinetic resolutions," *J. Am. Chem. Soc.*, vol. 139, no. 36, pp. 12441–12449, 2017.

[44] Malkov, A. V., Mariani, A., MacDougall, K. N., and Kočovský, P., "Role of noncovalent interactions in the enantioselective reduction of aromatic ketimines with trichlorosilane," *Organic Letters*, vol. 6, no. 13, pp. 2253–2256, 2004.

[45] Marshall, M. S., Burns, L. A., and Sherrill, C. D., "Basis set convergence of the coupled-cluster correction, $\delta_{MP2}^{CCSD(T)}$: Best practices for benchmarking noncovalent interactions and the attendant revision of the S22, NBC10, HBC6, and HSG databases," *J. Chem. Phys.*, vol. 135, p. 194102, 2011.

[46] Møller, C. and Plesset, M. S., "Note on an approximation treatment for many-electron systems," *Phys. Rev.*, vol. 46, pp. 618–622, 1934.

[47] Nishio, M., Hirota, M., and Umezawa, Y., *The CH/π Interaction*. New York: Wiley-VCH, 1998.

[48] Papajak, E., Zheng, J., Xu, X., Leverentz, H. R., and Truhlar, D. G., "Perspectives on basis sets beautiful: Seasonal plantings of diffuse basis functions," *J. Chem. Theory Comput.*, vol. 7, pp. 3027–3034, 2011.

[49] Parker, T. M., Burns, L. A., Parrish, R. M., Ryno, A. G., and Sherrill, C. D., "Levels of symmetry adapted perturbation theory (SAPT). I. Efficiency and performance for interaction energies," *J. Chem. Phys.*, vol. 140, p. 094106, 2014.

[50] Parrish, R. M., Parker, T. M., and Sherrill, C. D., "Chemical assignment of symmetry-adapted perturbation theory interaction energy components: The functional-group sapt partition," *J. Chem. Theory Comput.*, vol. 10, pp. 4417–4431, 2014.

[51] Parrish, R. M. and Sherrill, C. D., "Spatial assignment of symmetry adapted perturbation theory interaction energy components: The atomic SAPT partition," *J. Chem. Phys.*, vol. 141, p. 044115, 2014.

[52] Parrish, R. M., Sitkoff, D. F., Cheney, D. L., and Sherrill, C. D., "The surprising importance of peptide bond contacts in drug-protein interactions," *Chem. Eur. J.*, vol. 23, pp. 7887–7890, 2017.

[53] Pipek, J. and Mezey, P. G., "A fast intrinsic localization procedure applicable for ab initio and semiempirical linear combination of atomic orbital wave-functions," *J. Chem. Phys.*, vol. 90, no. 9, pp. 4916–4926, 1989.

[54] Pulay, P. and Saebø, S., "Orbital-invariant formulation and second-order gradient evaluation in Møller-Plesset perturbation theory," *Theor. Chim. Acta*, vol. 69, pp. 357–368, 1986.

[55] Purvis, G. D. and Bartlett, R. J., "A full coupled-cluster singles and doubles model: The inclusion of disconnected triples," *J. Chem. Phys.*, vol. 76, pp. 1910–1918, 1982.

[56] Raghavachari, K., Trucks, G. W., Pople, J. A., and Head-Gordon, M., "A 5th-order perturbation comparison of electron correlation theories," *Chem. Phys. Lett.*, vol. 157, pp. 479–483, 1989.

[57] Raynal, M., Ballester, P., Vidal-Ferran, A., and van Leeuwen, P. W. N. M., "Supramolecular catalysis. part 1: non-covalent interactions as a tool for building and modifying homogeneous catalysts," *Chem. Soc. Rev.*, vol. 43, pp. 1660–1733, 2014.

[58] Reddi, Y. and Sunoj, R. B., "Origin of stereoselectivity in cooperative asymmetric catalysis involving n-heterocyclic carbenes and lewis acids toward the synthesis of spirooxindole lactone," vol. 7, no. 1, pp. 530–537, 2017.

[59] Ringer, A. L., Figgs, M. S., Sinnokrot, M. O., and Sherrill, C. D., "Aliphatic C–H/π interactions: Methane-benzene, methane-phenol, and methane-indole complexes," *J. Phys. Chem. A*, vol. 110, pp. 10822–10828, 2006.

[60] Riplinger, C., Sandhoefer, B., Hansen, A., and Neese, F., "Natural triple excitations in local coupled cluster calculations with pair natural orbitals," *J. Chem. Phys.*, vol. 139, p. 134101, 2013.

[61] Schrödinger, LLC, "The PyMOL molecular graphics system, version 1.8." November 2015.

[62] Schutz, M., Hetzer, G., and Werner, H.-J., "Low-order scaling local electron correlation methods. i. linear scaling local mp2," *J. Chem. Phys.*, vol. 111, pp. 5691–5705, 1999.

[63] Schütz, M. and Manby, F. R., "Linear scaling local coupled cluster theory with density fitting. Part i: 4-external integrals," *Phys. Chem. Chem. Phys.*, vol. 5, pp. 3349–3358, 2003.

[64] Schütz, M., Werner, H. J., Lindh, R., and Manby, F. R., "Analytical energy gradients for local second-order Møller-Plesset perturbation theory using density fitting approximations," *J. Chem. Phys.*, vol. 121, pp. 737–750, 2004.

[65] Scuseria, G. E., Janssen, C. L., and Schaefer, H. F., "An efficient reformulation of the closed-shell coupled cluster single and double excitation (CCSD) equations," *J. Chem. Phys.*, vol. 89, p. 7382, 1988.

[66] Seguin, T. J. and Wheeler, S. E., "Competing noncovalent interactions control the stereoselectivity of chiral phosphoric acid catalyzed ring openings of 3-substituted oxetanes," vol. 6, no. 10, pp. 7222–7228, 2016.

[67] Seguin, T. J. and Wheeler, S. E., "Stacking and electrostatic interactions drive the stereoselectivity of silylium-ion asymmetric counteranion-directed catalysis," *Angew. Chem. Int. Ed.*, vol. 55, no. 51, pp. 15889–15893, 2016.

[68] Shao, Y., Gan, Z., Epifanovsky, E., Gilbert, A. T. B., Wormit, M., Kussmann, J., Lange, A. W., Behn, A., Deng, J., Feng, X., Ghosh, D., Goldey, M., Horn, P. R., Jacobson, L. D., Kaliman, I., Khaliullin, R. Z., Kus, T., Landau, A., Liu, J., Proynov, E. I., Rhee, Y. M., Richard, R. M., Rohrdanz, M. A., Steele, R. P., Sundstrom, E. J., Woodcock, H. L., Zimmerman, P. M., Zuev, D., Albrecht, B., Alguire, E., Austin, B., Beran, G. J. O., Bernard, Y. A., Berquist, E., Brandhorst, K., Bravaya, K. B., Brown, S. T., Casanova, D., Chang, C., Chen, Y., Chien, S. H., Closser, K. D., Crittenden, D. L., Diedenhofen, M., DiStasio, R. A., Do, H., Dutoi, A. D., Edgar, R. G., Fatehi, S., Fusti-Molnar, L., Ghysels, A., Golubeva-Zadorozhnaya, A., Gomes, J., Hanson-Heine, M. W. D., Harbach, P. H. P., Hauser, A. W., Hohenstein, E. G., Holden, Z. C., Jagau, T., Ji, H., Kaduk, B., Khistyaev, K., Kim, J., Kim, J., King, R. A., Klunzinger, P., Kosenkov, D., Kowalczyk, T., Krauter, C. M., Lao, K. U., Laurent, A. D., Lawler, K. V., Levchenko, S. V., and Lin, C. Y., "Advances in molecular quantum chemistry contained in the q-chem 4 program package," *Mol. Phys.*, vol. 113, pp. 184–215, 2015.

[69] Sharkas, K., Toulouse, J., Maschio, L., and Civalleri, B., "Double-hybrid density-functional theory applied to molecular crystals," *J. Chem. Phys.*, vol. 141, p. 044105, 2014.

[70] Sherrill, C. D., "Energy component analysis of π interactions," *Acc. Chem. Res.*, vol. 46, pp. 1020–1028, 2013.

[71] Stone, A. J., *The Theory of Intermolecular Forces*. Oxford: Oxford University Press, 1996.

[72] Szabo, A. and Ostlund, N. S., *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*. New York: McGraw-Hill, 1989.

[73] Szalewicz, K., "Symmetry-adapted perturbation theory of intermolecular forces," *WIREs Comput. Mol. Sci.*, vol. 2, pp. 254–272, 2012.

[74] Turney, J. M., Simmonett, A. C., Parrish, R. M., Hohenstein, E. G., Evangelista, F. A., Fermann, J. T., Mintz, B. J., Burns, L. A., Wilke, J. J., Abrams, M. L., Russ, N. J., Leininger, M. L., Janssen, C. L., Seidl, E. T., Allen, W. D., Schaefer, H. F., King, R. A., Valeev, E. F., Sherrill, C. D., and Crawford, T. D., "Psi4: An open-source *ab initio* electronic structure program," *WIREs Comput. Mol. Sci.*, vol. 2, pp. 556–565, 2012.

[75] Uyeda, C. and Jacobsen, E. N., "Transition-state charge stabilization through multiple non-covalent interactions in the guanidinium-catalyzed enantioselective claisen rearrangement," *Journal of the American Chemical Society*, vol. 133, no. 13, pp. 5062–5075, 2011.

[76] Weigend, F., "A fully direct ri-hf algorithm: Implementation, optimized auxiliary basis sets, demonstration of accuracy and efficiency," *Phys. Chem. Chem. Phys.*, vol. 4, pp. 4285–4291, 2002.

[77] Weigend, F., "Accurate coulomb-fitting basis sets for h to rn," *Phys. Chem. Chem. Phys.*, vol. 8, pp. 1057–1065, 2006.

[78] Weigend, F., Kattannek, M., and Ahlrichs, R., "Approximated electron repulsion integrals: Cholesky decomposition versus resolution of the identity methods," *J. Chem. Phys.*, vol. 130, p. 164106, 2009.

[79] Weigend, F., Köhn, A., and Hättig, C., "Efficient use of the correlation consistent basis sets in resolution of the identity mp2 calculations," *J. Chem. Phys.*, vol. 116, pp. 3175–3183, 2002.

[80] Werner, H.-J. and Manby, F. R., "Explicitly correlated second-order perturbation theory using density fitting and local approximations," *J. Chem. Phys.*, vol. 124, p. 054114, 2006.

[81] Werner, H.-J., Manby, F. R., and Knowles, P. J., "Fast linear scaling second-order Møller-Plesset perturbation theory (MP2) using local and density fitting approximations," *J. Chem. Phys.*, vol. 118, no. 18, pp. 8149–8160, 2003.

[82] Wheeler, S. E., Seguin, T. J., Guan, Y., and Doney, A. C., "Noncovalent interactions in organocatalysis and the prospect of computational catalyst design," *Acc. Chem. Res.*, vol. 49, no. 5, pp. 1061–1069, 2016.

VITA

Brandon W. Bakr was born in 1991 in Dothan, Alabama to A. D. Bakr and Merlene Bakr. He graduated from Dothan High School in Dothan, Alabama in 2009. He then went on to Auburn University in Auburn, Alabama where he worked under the guidance of Dr. Andras Bezdek and Dr. Konrad Patkowski. He graduated *cum laude* with honors and a major in Chemistry and a minor in Mathematics. He then did graduate studies in theoretical chemistry under the guidance of Dr. C. David Sherrill at Georgia Tech.